Population Genetics of Western Mediterranean Islands - Malta: a case study

A thesis submitted to The University of Manchester for the degree of

Doctor of Philosophy

in the Faculty of Life Sciences

2012

Josef Caruana

List of Contents

Word count: 49,497

List of figures:

List of Tables

List of Abbreviations

aDNA	Ancient DNA
AMOVA	Analysis of molecular variance
DMF	N,N-dimethylformamide
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
HVR1	Hypervariable region 1
HVR2	Hypervariable region 2
IPTG	Isopropyl-beta-D-thiogalactopyranoside
LBA	Luria-Bertoni agar
mtDNA	Mitochondrial DNA
PC	Principal Component
PCA	Principal Component Analysis
PAGE	Polyacrylamide gel electrophoresis
SDS	Sodium dodecyle sulfate
SNP	Single Nucleotide Polymorphism
STR	Short tandem repeat
TBE	Tris/Borate/EDTA buffer
Y-STR	Y-chromosomal short tandem repeat

Abstract

The University of Manchester

Josef Caruana

Doctor of Philosophy

Population Genetics in Western Mediterranean islands – Malta: a case study

28/10/2012

In order to gain a greater understanding of the genetic makeup of the Maltese population, mitochondrial DNA HVR1 and HVR2, and Y-chromosomal and autosomal STRs were amplified in a representative sample of the Maltese population. The results showed that the Maltese have close genetic ties with Sicily and mainland Italy both from a matrilineal and a patrilineal perspective, whilst no conclusive evidence was found for a Phoenician link between the Maltese and the Lebanese population. In order to try and gain an insight into the Maltese population throughout time, a study was conducted on three Maltese archaeological burial places dating from the Neolithic to the Roman period. The study extracted and amplified ancient DNA sequences from these three sites and compared the resulting mtDNA sequences with the modern Maltese population. The results showed that aDNA survives in the Maltese archaeological record, and that some haplotypes found during the Roman period in Malta are also found in the modern day population, whilst other haplotypes present in the archaeological samples are not visible in the modern Maltese population.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on Presentation of Theses

Dedication

I would like to dedicate this work to the island nation of Malta and its people. May this work help in bringing more unity and a sense of identity and pride amongst its population.

Acknowledgments

I would like to thank my supervisor Terry Brown, without whose help and guidance this work would have never been conducted. I would also like to thank Keri Brown for helping me with the sampling of the Maltese inhumations and giving me the foundations of aDNA analysis. Special thanks go to Abigail Bouwman, Sarah Ayling and Sandra Bunning, who had the patience to teach me techniques I was unfamiliar with in the laboratory and in data analysis. Thanks also goes to Claudia Grimaldo, who helped me with the modern STR genotyping, showing me what she had achieved using the same method on plant genomes. A thank you must also go to Naglaa Abu-Mandil and Romy Muller for their support in the laboratory.

I must not forget to mention the people who have been of great help to me outside the laboratory. These are Laura Robinson who encouraged me constantly and had to suffer my endless rants when things failed in the laboratory. Stephanie Fisher, Duncan Ayers and Shail Segobin who were with me throughout the work. They were always a shoulder to cry on or a friend to celebrate with, according to the circumstances. Other people from my years in Manchester also deserve thanks especially the security guards at Opal Gardens. I must also thank my parents and my brother, without whose support, both financial and moral, this work would have never started, let alone come to a conclusion.

Special thanks also go to the Superintendence of Cultural Heritage and the Malta Biobank, for providing the ancient and the modern samples respectively, without their help the genetic knowledge about the Maltese population would be poorer.

Chapter 1: Introduction

1.1: Maltese archaeology

Malta is a small island situated at the centre of the Mediterranean, 60 miles south of Sicily. Malta enjoys a varied and rich history and prehistory (Table 1). The reason for this is that it was occupied and colonised by almost all the Mediterranean powers that were present throughout history. Even when this did not happen, such as during prehistory, a link with the outside world was always maintained, even though sometimes this link is less evident than at other times.

Key Events	**Estimated Dates**
Initial peopling of the Maltese islands from Sicily	5000B.C.-3600B.C.
Period of increasing complexity in the architecture of the island. Tarxien period reaches the apex of this complexity. Kercem samples date to the Tarxien period.	3600B.C.-2500B.C.
Change in material culture resulting in the Bronze Age period	2000B.C.-800B.C.
Phoenician and Carthaginian settlement of the island	800B.C.-218B.C.
Roman period commencing after the conquest of Malta in the Second Punic War. Tal-Gardina and Tal-Barrani samples date to the last centuries of the Roman period.	218B.C.-395A.D.
Barbarian invasions of the Roman Empire including the probable conquest of Malta. No knowledge of what	Around 450A.D.

happened in Malta during this period.	
Byzantine period where Malta was under the Eastern Mediterranean sphere of influence.	500A.D.-870A.D.
Arab conquests in the Mediterranean include Malta and Sicily.	870A.D.-1090.A.D
Norman conquest of Sicily and Malta, this is the period when Malta returns to the Western Mediterranean sphere of influence.	1090A.D.-1265A.D.
Angevins acquire the island through inheritance, but lose it together with Sicily in the Sicilian Vespers revolt.	1266A.D.-1282A.D.
Aragonese period where little is known about life on the Maltese islands at the time.	1282A.D.-1414A.D.
Unification with the Spanish crown, including periodic revolts of the Maltese against their feudal lords, and attacks by corsairs from the Barbary coast.	1414A.D.-1530A.D.
Malta given to the Knight Hospitallers of Saint John. Steady population growth is observed after the Great Siege of 1565 where a possible population bottleneck occurred. Malta also becomes a cosmopolitan centre.	1530A.D.-1798A.D.
Brief period of Napoleonic French occupation, ending with the insurrection of the Maltese population.	1798A.D.-1800A.D.
The British come to Malta at the request of the Maltese to help with the insurrection, and ruled the country as a special protectorate, after which the island become a crown colony	1800A.D.-1964A.D.

Independence granted in 1964, after which Malta became a republic in 1974	1964-present

Table 1: Different periods of Maltese prehistory and history.

In his book Trump (2002) states that the future of Maltese prehistory looks bright since innovative techniques to study the archaeological remains are constantly being used. Trump mentions the study of ancient DNA (aDNA) as a technique which might yield very interesting results about Maltese prehistory. In this project one of the aims of the study is to see how feasible it is to study Maltese archaeological inhumations by using aDNA techniques. An attempt will also be made to gather as much information as possible from the remains of three particular sites in order to help archaeologists interpret the sites better, and also to shed light on the implications the results obtained have on Maltese history in general.

The archaeological story of Malta starts in the Neolithic with the arrival of people on the island. Due to similarities in pottery design and it being the closest land, Sicily is thought to be the place from where these people originated (Trump 2002). Whilst the first signs of habitation in Malta were found in caves and then in small structures, throughout the rest of the Neolithic the inhabitants of the island progressed in the building of worshipping and funerary buildings of increasingly complex structure. This period is called the Temple period and it reached its apex in the complex structures constructed during the Tarxien phase, both above ground as in the case of the Tarxien temples, and even below ground as in the case of the Hal-Saflieni hypogeum. This expansive and laborious process of building increasingly complex structures came to an abrupt end just before the Bronze Age.

At the start of the Bronze Age period the Maltese islands seemed to have suffered some kind of change. Whether the change seen in the material culture was due to a cultural change in the residents or whether this happened with the arrival of new people on the island is still debatable (Trump 2002). The change meant that structures suddenly changed into more simplistic forms and the burial method changed from inhumation to cremation. The dawn of Maltese history is considered to have occurred with the arrival of the Phoenicians who brought with them their alphabet at around 700 B.C. Some evidence suggests that the Phoenicians coexisted with the previous inhabitants of the island for some time before any trace of the previous inhabitants disappears from the archaeological record The Phoenician presence in Malta lasted for centuries and this is evident in the archaeological record with the remains of the temple at Tas-Silg, and with the several burial chambers that are found scattered all over the islands dating from that period.

During the Second Punic War, Malta was taken over by the Romans (Livy xxi.51). During the Roman period archaeological evidence suggests that there was a thriving population in Malta, with the ancient city of Mdina at the centre of the island covering a greater area than it does today. Further evidence of the thriving population is the remains of the Roman villas that are found on the island. Malta was also being used as a place where olives were grown and pressed. This fact can be seen from the remains of the Roman complex at Zejtun, where a large underground chamber used to store olive oil was found. Evidence that there was extensive contact with the rest of the Roman Empire can be seen from the archaeological record. A tombstone was found which mentions that the person that had been buried within was a travelling actor from Sicily (Bonanno 1992). When a Roman governor called Verres stole sacred artefacts from the temple at Tas-Silg, Cicero accused Verres of it, whilst at the

same time describing the sanctity of the temple, and that Malta had a weaving industry (Cicero Book 4 XLVI).

Cicero also considered exiling himself in Malta in one of the letters he sent to a friend, although we do not know if he actually did this. Other important evidence which shows that Roman Malta was well informed with what was happening in the empire at the time are the busts which were found at the Roman domus in Rabat. These busts portray the imperial family (Bonanno 1992). All of this contact with the outside world implies that there is a great probability that new inhabitants were settling in Malta at the time. This might be true especially in the Imperial era, where the Maltese islands were in the middle of mare nostrum and not at the outskirts of the Roman Empire, as they were until the Third Punic War.

A question I will try to tackle in my study about this period is the belief that has been passed on by generations in Malta, that the Maltese are descendents from the Phoenicians. This idea has been proposed by politicians throughout time and also by some ancient scholars who also argue that the basis of the Maltese language is Phoenician (Abela 1647). The proposal is contested by the majority of present day scholars who argue that the modern Maltese language has Semitic roots, but that these were of Arabic origin at the start of the Middle Ages and not of earlier origin (Brincat 1991, Mifsud 1995). My aim is to try and determine through modern population genetics if a relevant link can be made between the people from modern Lebanon and the present day Maltese population. A link between the Maltese and the Lebanese populations would imply that the Maltese have still got Phoenician blood in them, as the Phoenicians were the last common link the Maltese had with the Lebanese, since all successive colonisations came from the north or the south, but never from the Near East.

The fact that Maltese coinage under the Roman Empire first appeared with Greek writing on it, and only later with Latin inscriptions, shows that the Maltese kept close links with Sicily, which was still heavily influenced by Greek culture due to the Greek colonies that had been set up there (Bonanno 1992). We do not know to what extent the population changed in Malta through these times, and if the island saw an influx of new arrivals from the Roman Empire, or if the change in culture was conducted by the people who were already residing in Malta.

After the fall of the Roman Empire the island is thought to have been occupied by the northern barbarians that had been the cause of the downfall of Rome (Vella 1975), and subsequently then by the Byzantines. There is a great ongoing debate whether the Byzantines abandoned Malta, or if it was taken from them by the Arabs (Wettinger 1986). This point is very interesting in terms of the Maltese gene pool, and an attempt will be made to determine genetically if the island was really deserted for a period during this time.

The Arab period of the Maltese islands left cultural traces which are visible to this day. These are seen in the Maltese language, were about 60% of the vocabulary has a Semitic origin, and also in the architecture of the islands. Whilst in recent years distinctive architectural features such as the Arabic peep hole in houses has disappeared, other things such as the flat roofs and the plan of a traditional Maltese farmhouse have got a distinctive North African flavour to them (Luttrell 1975). During this period it is thought that the whole island turned Muslim; this is implied due to the lack of Christian objects and places of worship, which are very dominant both before and after this period. This does not mean that the inhabitants of the islands were treated as equal to their masters; this was seen in the

records of a Byzantine siege of the islands in which the Arab elite agreed to start treating the population as equals if they helped them in the siege (Dalli 2006). The local population complied and achieved a greater standing in society.

After the Arab period, the Normans came to the island, and from that time till the arrival of the British on the island, Malta started having closer ties with the European mainland. During the Medieval period and the period of the Knights of Saint John, there were several times when the population of the island suffered drastic reductions in size due to either war or pestilence, an example of such was the removal of almost the whole population from Malta's sister island Gozo in 1551; they were all sold into slavery by Turgut Reis, a pirate from the Barbary coast (Vella 1975). Another attack from such pirates occurred in medieval times when a third of the population was lost during a siege that they mounted on the Maltese main city of Mdina (Dalli 2006). The arrival of the Knights of Saint John also brought in itself a change in the Maltese population, this happened with the advent of the Rhodian refugees which left their birthplace in Rhodes and followed the Knights of Saint John into Malta (Vella 1975). In my study of Maltese population genetics I will attempt to see if there are any traces of these migrations and emigrations and possible bottlenecks which can be detected by looking at the Maltese gene pool.

After a very brief interregnum of French occupation and a Maltese revolt, the Maltese islands then passed under the British for 164 years, in which Malta became the home base of the Mediterranean fleet. This meant that thousands of sailors and soldiers from all over the commonwealth used to stay in Malta for long periods of time. This long standing connection with the Empire can still be seen to this day with the English names of children, English street names, and English still being an official language of Malta together with Maltese.

This means that over time a lot of people from different parts of the Mediterranean lived in the Maltese archipelago, and one of the questions this study will try to answer is how many of them left a visible genetic imprint on the Maltese population.

1.2: The archaeological sites which were sampled

1.2.1: The Kercem site

The Kercem site is the only site studied which is from Gozo, Malta's sister island (Figure 1). The site was discovered in 2008 and is the most recently excavated site. When it was discovered the site was not immediately excavated, but since the site had experienced flooding, it was left alone for a year until the site had dried up after which the site was excavated.

Figure 1: Location of the Tal-Barrani, Kercem and the Tal-Gardina sites on the Maltese Islands. The red circle on the inset map indicates the place of Malta in the middle of the Mediterranean (After: http://maps.google.com).

The site dates to the late Tarxien phase (3,000 - 2,500B.C.) which is just before the supposed break which happens in Maltese prehistory. It is also situated just a few kilometres away from the large funerary complex of the Xaghra Stone Circle which dates to the same period but contained thousands of inhumations in it. The difference between the two sites is that whilst the inhumations in Kercem were buried one above the other (Figure 2), the ones in Xaghra had their bones piled up together in groups after the body had decomposed (Malone *et al*. 2009). The site which was excavated was a tomb which had three inhumations buried in. The site was unique for the Maltese island as the inhumations had a stone slab between each of them which had never been seen before in Maltese archaeology (Anthony Pace 2001, personal communication). The site is thought to have been part of a larger funerary area, but the rest of the tombs were lost during excavation work which was done for constructing buildings in the vicinity. The site revealed some Tarxien phase pottery but no other things were buried with the bodies. The archaeologists were not sure if the three persons buried in the same tomb meant that the persons were related (Anthony Pace 2001, personal communication).

Figure 2: The inhumations found at Kercem whilst being excavated.
(From: http://www.timesofMalta.com/articles/view/20110108/opinion/protecting-the-most-significant-buildings-monuments-and-features-of-Maltese-islands-55.344349).

When the samples were taken for DNA testing, the archaeologists were warned that due to the site having been flooded and its age, the chance of the samples yielding aDNA was lower than in the other sites which were sampled. The site would serve as a benchmark as to whether ancient archaeological remains which had not been in optimal conditions would still yield results in the Maltese islands.

1.2.2: The Tal-Barrani and the Tal-Gardina sites

The two other sites which were studied were the Tal-Barrani and the Tal-Gardina funerary complexes. The sites are found in different contexts. The Tal-Gardina site is situated just outside the ancient Roman city walls, and thus may be considered as a burial place in an

urban context. The Tal-Barrani site is found on the south of the island in a rural context, with the nearest industries we know about in Roman times near it being a country house with an oil press and a large oil storage area underneath it, situated two kilometres away. Another site which was situated slightly further away was the port of Marsa, which archaeological evidence suggests was used as one of the main ports of the island during Roman times. Both sites date to the 4th century A.D.

The Tal-Barrani Catacombs is a site which was discovered by accident when new road works were being done. The site was excavated seventeen years ago and in it two features where found; feature one can be seen in Figures 3A and 3B. Feature one was a complex where six skeletons had been interred and these were found to be articulated, whilst the other feature contained a variety of parts from different skeletons. The report from this site has not yet been published as further investigations are still under way, and it is hoped that findings from this study will also be included in the site report. When analysed by an osteologist the skeletons from this site showed thickening of the bone, which indicates that the persons buried had been used to a high level of physical labour during their lifetimes. The six articulated skeletons that made up feature one have all been sampled. From the material culture found on the Tal-Barrani site the period in which this site was in use is thought to be the late Roman period (300-400A.D.).

Figure 3A and 3B: Feature one and entrance to the Tal-Barrani site (From http://www.angelfire.com/ma/cop994/2index.html).

The Tal-Gardina site is located just outside the ancient Roman city walls. It was discovered in late 2007. A plan of the site was made available to the researcher (Figure 4). It is thought that the site might contain up to sixty different inhumations, but unfortunately only a few of the inhumations found until now were articulated. There are several reasons for this. According to the osteologist one of the reasons could be the removal of bodies in order to make place for other bodies. Another reason could be the gases that decomposition produces, which could make the bodies fall over from the loculus where they would have been laid. Another reason for disarticulation proposed by the osteologist is the presence of animals which might have moved the bones around.

The site was used at the same time as the Tal-Barrani site, but the bones recovered from it are thinner; this indicates that the persons buried in this site were not used to as much physical labour as those of Tal-Barrani. Nine samples were taken from the remains of the site. Six of the samples were taken from three articulated skeletons, whilst another three samples were taken from three right tibias which were found in very close proximity to each other onsite. The osteologist working on the site is also still studying the bones which are still being uncovered.

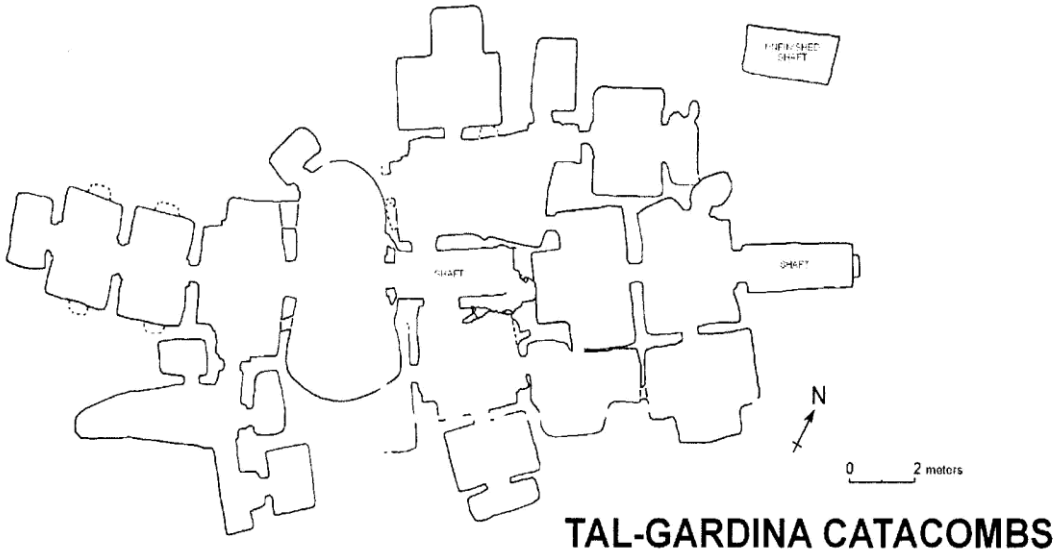

Figure 4: Site plan for the Tal-Gardina Catacombs (After Vella and Vella-Osanich 2008).

1.3: Ancient DNA

1.3.1: The study of ancient DNA and its authentication

The study of aDNA is a relatively new discipline which has only been in existence for the last 25 years (Willerslev and Cooper 2005). Whilst at first it was thought that this new discipline was producing great results, the researchers soon realised that a lot of the results being obtained were not truly aDNA but modern contamination. Thus the need arose for a set of rules to be brought in, in order to ensure that there was consistency and reliability throughout the field, and over the years specific modes of practice were developed in order to ensure that the results are as accurate as possible (Cooper and Poinar 2000).

One can never be completely sure that what has been amplified during a PCR reaction is ancient DNA, but certain safeguards can be taken in order to arrive at a point where the most

logical explanation for the DNA sequence is that it is truly ancient DNA. There are several ways to do this and different labs use different methods, but the most common established protocols for DNA authentication are the following.

The biomaterial from which the DNA is going to be extracted needs to be assessed in aDNA studies. The reason is that good preservation in bone and teeth is an indicator that aDNA might still be present in the material. This assessment can be done by both physical observation of the material and also by conducting tests in order to determine the survivability of other material associated with DNA survivability, such as by analyzing the amount of collagen still present in the sample. A rough indicator of collagen still being present can be observed when one is taking samples from a bone with a hacksaw or a drill. If the smell of burned hair is observed during the sampling procedure, it is a good indication that collagen is still present.

Another way used to gain an insight into whether the amplified DNA obtained is really aDNA is to look at the miscoding lesions that arise in ancient DNA. Type 2 transitions (cytosine to thymine and guanine to adenine) have been observed to increase in the 5\ and the 3\ ends of the strand. Thus this can be used in order to aid in the determination of whether a sequence is truly aDNA (Stoneking and Krause 2011).

Whilst the contamination from intact DNA can be removed by standard cleaning methods (Kirsanow & Burger 2011), it is much harder to remove degraded contamination which might be already present in the samples or have been introduced in the laboratory. This degraded contamination might also outcompete the authentic aDNA present (Kirsanow &

Burger 2011) and thus areas in which this contamination is known to be more concentrated such as in the outer 2mm of the bone must be removed (Bouwman *et al.* 2006).

Next Generation Sequencing is very useful in aDNA studies as it is able to target DNA fragments of less than 80bp which cannot normally be amplified with classical PCR methods. Fragments less than 80bp long under normal circumstances contain a much higher proportion of aDNA sequences than contamination and thus the contamination should not be able to outcompete the aDNA (Kirsanow & Burger 2011).

Care in protocols must always be taken in order to make sure that the decontamination procedures used do not degrade and fragment the DNA into smaller amplifiable pieces, as this could lead to the amplification of modern contamination which can than outcompete aDNA (Kirsanow & Burger 2011).

A way in order to help authenticate aDNA is to design multiplexes where whilst the rest of the primer sets are designed to amplify DNA fragments of less than 150bp, one of the primer sets is designed to yield an amplification of over 500bp (Abigail Bouwman personal communication). The reason is that if this primer set amplifies DNA of such a range it can normally be safely assumed that the sample is contaminated unless it was found in an exceptional preservation status as such DNA fragments are rare (Cooper & Poinar 2000). The reason is that DNA would degrade in shorter fragments than 500bp over several centuries or millennia.

All of these authentication aids must be used in addition to the main standard operating procedures when dealing with aDNA studies. These procedures are designed to minimize

the chance of contamination and have been adopted almost universally in aDNA studies. These procedures are as described by Cooper and Poinar (2000):

That no ancient DNA work is conducted were modern DNA work has been conducted, aDNA work should be done in specially designed laboratories physically separate from were modern DNA work is conducted. The places were bone powdering, extraction and the amplification reactions are prepared should be physically separate from each other and from the modern DNA laboratory. This is done in order to minimize the risk of cross-contamination.

Protective clothing must be worn and decontamination procedures must also be set in place and followed rigorously in these rooms. This is done in order to minimize the risk of contamination and also to ensure that the laboratories are treated the same way by different laboratory members at different times who might be using them.

Control amplifications must also be carried out in each step of the procedure. This should help in identifying contamination and also try and determine at what time the contamination took place.

Results should be as far as possible repeated both on the same sample and also on different samples from the same inhumation in order to ensure consistency. Cloning should also be performed in order to ensure that contamination in an amplification is identified by different sequences being visible, something which may not always be possible with direct sequencing due to the contaminant overwhelming the signal.

Wherever possible all the results obtained should be replicated by an independent laboratory. Although this might not always be possible due to cost considerations, such a step strengthens the claim of authenticity of aDNA.

1.3.2: DNA degradation

A very important factor to consider when working with aDNA is the fact that aDNA is constantly degrading. This is because the repair mechanisms that maintain the DNA under normal circumstances stop working as soon as the organism dies. Thus when a researcher is selecting samples from which to extract and analyse aDNA, the researcher must be careful that the samples have been subjected to conditions that enhance the likelihood of DNA survival (Wayne *et al*. 1999). One way to look at the likelihood of DNA being present in a sample is by estimating the thermal age of the sample in question (Smith *et al.* 2003). The most suitable conditions for the preservation of aDNA in bone are when the specimen has been subjected to a cold and dry environment and the sample is less than 100,000 years old (Wayne *et al.* 1999). This is shown in a study conducted by Höss *et al.* (1996), where samples of DNA coming from the Arctic region survived much better than the DNA originating from samples found in hot countries.

1.4: Burial and storage conditions

Burger *et al.* (1999) investigated how environmental factors affect the quantity and quality of DNA that can be retrieved from teeth which were of similar age but that were in different conditions. The results showed that constant low temperature yielded good reproducible

DNA results even though it was in a wet environment with microbiological activity observed in the bone (61-90% successful amplification with 40-91% reproducibility). Samples which were in dry conditions and with little microbiological activity but with a constant high temperature showed only a 35% DNA amplification success ratio and of the successful amplifications only 12% could be reproduced. The authors say that this shows that the temperature the specimens are situated in when found are more of a factor than the microorganism activity the samples might have suffered, even though microorganism activity in the samples does degrade the quality and quantity of DNA that can be obtained. This study shows that whilst as Wayne *et al.* (1999) aDNA is best preserved in cold and dry conditions, it is better to have a cold and wet environment than a dry and hot environment for aDNA preservation.

When humidity was researched in the same study it was found that whilst high humidity does not discernibly lower the amount of DNA in a sample, it reduces the DNA quality drastically meaning that the results become less reproducible.

The results of a study of post excavation storage before DNA work was conducted are also very important to consider. The results in the Burger *et al.* (1999) study showed that whilst samples which were stored for 16 years at room temperature produced roughly the same amount of DNA amplification products as the ones which were stored at -20^{o}C, the reproducibility of the PCRs fell dramatically from 86% to 40%. This decrease in reproducibility was also seen with the DNA extracted from 0.1g of a tooth sample after 2 years at room temperature.

1.5: Thermal age

Thermal age differs from the actual age of the sample in that it is the theoretical age of the sample calculated from the degradation suffered due to the temperature at which it has been held (Smith *et al.* 2003). The thermal age model whilst indicative of the extent of DNA degradation that may be expected from a sample is not always reliable. This is shown from the work conducted by Poinar *et al.* (2003) who managed to extract DNA from sloth coprolites which were four times older than the theoretical thermal age limit of DNA recovery.

Since the Maltese specimens have been subjected to higher temperatures than would be optimal due to the Maltese hot dry summers and short mildly cold wet winters, their thermal ages are increased considerably. This will impinge on the samples from the Tal-Barrani and Tal-Gardina catacombs, but the effect should be less than for the Kercem site since they hail from the historical era and not prehistory and thus are over 2,000 years younger than the Kercem samples. The Kercem site on the other hand is more problematic due to the age of the site and the fact that it has been flooded, with microorganism activity also present on the bone surface. Unfortunately most of the sites being studied have now been sealed and so the temperature could not be taken, thus determining the exact thermal age of the samples was not possible.

Water is one of DNA's worst enemies. Water attacks the DNA, breaking the β-N-glycosidic bond, removing a base from the polynucleotide which might eventually break the strand in two (Brown and Brown 2011). Thus samples which have been exposed to water will suffer more rapid DNA degradation which results in less aDNA being amplified. Because of this,

the known history of water exposure to sites was noted. And whilst Tal-Gardina and Tal-Barrani only seem to have suffered periodic rain exposure, the site in Kercem was flooded more often and thus the chance of DNA amplification from the site was reduced.

1.6: Ancient DNA contamination

The study of aDNA is a relatively new science. The first papers which were brought out about the amplification of DNA from archaeological material caused quite a stir in scientific circles. After more research it was discovered that most of the initial aDNA studies had not really found aDNA but contamination. A lot of faith in the field was lost after this discovery, and in order for contamination not to be mistaken for authentic results again, researchers started taking more precautions (Cooper and Poinar 2000).

Contamination can occur from various sources; it can come from the archaeologists who excavated the site, the osteologist who studied the biological material, and the curator who was entrusted with the inhumations and the general dust in the air of where the samples were stored. Another source of contamination is the researcher and the other people who work in the lab and use the same facilities in order to do their work. Another source of contamination is cross contamination between samples. DNA can also be present in the reagents used in such studies, and another possible source of contamination which must be considered is that the microfuge tubes, pipettes and other equipment used in the analysis might have already been contaminated with DNA from the manufacturers.

Studies have shown that in bone the majority of contamination is found in the outer 2mm of the bone (Bouwman *et al.* 2006). Due to this, precautions are taken when sampling and

extracting DNA in order to ensure that the risk of contamination is reduced as much as possible.

1.7: Genetic markers and their use in population genetics

In order to reconstruct the genetic past of a set of samples, one must be wary about what markers one is going to use. Thus Jorde *et al.* (1998) in his study explains that choosing polymorphisms which are selectively neutral will result in a more accurate picture of the genetic histories of populations, and the relationships between them.

1.8: Mitochondrial DNA (mtDNA)

MtDNA is useful as a genetic marker as it is found at a very high copy number when compared to nuclear DNA in the human cell. Satoh and Kuroiwa (1991) estimated that there are about 500 copies of mtDNA in most cells. This means that mtDNA can prove very useful when studying degraded DNA such as aDNA, since the probability of amplifying DNA is greater than in chromosomal DNA. The biggest advantage of mtDNA is that it is passed on matrilineally; this is an advantage when using mtDNA in order to study populations. By analysing single mutations that occurred in the mtDNA throughout history, one can infer with a certain degree of certainty events that occurred in the past due to the lack of recombination of the marker, such studies as the introduction of modern humans into Europe and the successive migrations that occurred (Richards *et al.* 1998) are a result of the properties of mtDNA. These studies are done by the use of a molecular clock in order to

infer divergence times. The molecular clock determines the rate at which DNA mutates in an organism throughout time.

One of the advantages of mtDNA is that it mutates at a higher rate than nuclear DNA (Vona *et al.* 2001). This is probably due to fewer repair mechanisms existing in the mitochondrion and the lack of proof-reading capabilities in the mtDNA polymerase which as a result increases mutations during replication (Butler 2005). This feature gives the researcher more precision in determining migrations which have occurred in only the last few tens of thousands years of human history. The ability of identifying different groups from their haplogroups means that questions about population movements can be studied by analyzing the geographic distribution of lineages on gene trees (Richards *et al.* 2001). A haplogroup is defined as one of the major sequence classes of mitochondrial DNA present in the human population (Brown and Brown 2011). Polymorphisms define these clades. Haplotypes here are defined as in Richards *et al.*(2000); where these are sequence types of hypervariable region one (HVR-I) sequences and position 73 of hypervariable region two (HVR-II), which might comprise several lineages (an individual subject's sequence), thus as explained by Richards *et al.*(2000) "a particular sequence type might comprise several lineages, if several individuals in a population sample display the same sequence type". Although this method of observing, describing and analyzing the distribution of genotypes and trying to apply past scenarios to them, which is called phylogeography (Kidd and Ritchie 2006) might not be foolproof, it is the best tool available to modern geneticists when studying evolution and population migrations throughout time.

Whilst restriction fragment length polymorphism analysis was at first used in mtDNA studies of human populations, as in the case of the discovery of the out of Africa origins of

humans (Cann *et al.* 1987), over time the sequencing of parts of the control region of the molecule became more popular. The reason was that hypervariable regions (HVR) 1 and 2 were more informative, since the control region has a higher mutation rate than the coding regions (Vigilant *et al.* 1991) and thus most studies if not both HVR started to at least include HVR1 in their analysis (Richards & Macaulay 2001).

1.9: Y chromosome

Just as single nucleotide polymorphisms in the mtDNA can be used in order to study migration through the female line, SNPs in the non-recombining region of the Y chromosome can be used in order to study population genetics on the patrilineal side, since the Y chromosome is passed on from father to son. The non-recombining region of the Y chromosome is used in population genetic studies in order to make sure that parts of the female DNA does not get involved in the analysis, thus rendering its flow from father to son not truly accurate, due to the mother's DNA also getting involved. When a locus that is analysed is non-recombining it makes sure that the locus has a single genealogical history (Richards & Macaulay 2001). Autosomal DNA suffers from recombination.

The discovery that the non-recombining region of the Y chromosome is more polymorphic than previously thought has allowed for worldwide trees to be drawn distinguished by different polymorphisms (Richards & Macaulay 2001).

These trees have given an insight into various questions about population migrations, an example is the study of Semino *et al*. (2004) which using a combination of SNP markers and

microsatellite loci showed how Y chromosome haplogroups E and J spread into the Mediterranean, with haplogroup E-M81 being present in Iberia and Sicily due to recent gene flow from North Africa whilst other haplogroups such as E-M78 showed a spread from the Balkans (Semino *et al.* 2004).

Another tool at the disposal of the geneticist when conducting population genetic studies are short tandem repeats (STRs), also referred to as microsatellites. These are small repeats of DNA consisting of two or more nucleotides repeating themselves for varying lengths in different individuals. The STRs normally used for studies are found in regions of DNA that are thought to be selectively neutral. These STRs can be used to ascertain relationships between individuals. This is achieved by mapping enough STR loci of an individual, and thus creating a genetic profile for the person, which is then compared to the other person in question (Butler 2005). The STRs used in forensic case studies and in population genetics are those found in either the Y and X chromosomes, or the autosomes. The larger the number of STRs that are used in a study, the more accurate the analysis derived from the study becomes (Butler 2005).

1.10: Analysing mtDNA and MSY sequences

The traditional way of analyzing mtDNA and Y-STR polymorphisms is by using gene trees in order to try and determine how different samples are related to one another. Various types of tree building methods can be employed which can give different results according to the method they use for building the trees and the assumptions made by the program. The final result is then displayed as a phylogenetic tree.

Some methods such as the neighbor-joining method use genetic distance in order to build the tree. This method starts by linking the samples which have got the closest distance to each other and continues building the tree upwards until all the samples have been placed in the tree. This method is similar to the Fitch-Margoliash method of tree building, but whilst the neighbour joining tree gives the same weight to each taxon, the Fitch-Margoliash method uses a least-square method and gives more weight to taxa which have a smaller genetic distance between them than to others.

The Fitch-Margoliash method uses the least squares method in order to cluster data. The method tries to position the closest related sequences together by calculating the two samples which have the least distance from each other in comparison to the average of the other sequences, the distance between different samples are then calculated using the same method, the furthest out the tree goes, the least related the sequences are. The advantage of the Fitch-Margoliash method is that this method allows for different rates of mutations from the root sequence when compared to other methods such as the Unweighted Pair-Group Method with Arithmetic mean (UPGMA) (Saitou and Imanishi 1989).

Maximum parsimony has also been used extensively in studying phylogenetic relationships in mtDNA. This method builds trees and assumes that that the simplest tree needing the least number of steps is the correct one. Some of the problems in using this method are that only informative sites of the sequence are used and that not all homoplasies that have occurred will always be shown. In this method of analysis a matrix of different attributes is used in order to determine phylogenetic trees according to how the different attributes are scored (Saitou and Imanishi 1989). After the phylogenetic trees are constructed the one which shows the least amount of mutations between the samples possible (and thus is the shortest

tree in terms of mutations) is selected. Sometimes more than one possible shortest tree is present in a set of data. Unfortunately the tree obtained does not necessarily mean that this is the correct evolutionary pathway. The reason for this is homoplasy were more than one mutation in the same site may be hidden over in the tree, especially if the mutation reverted back to the original over time. Torroni *et al.* (1994) used this method in order to identify different haplogroups for the mtDNA and the Y-chromosomal DNA of a sample of Native Americans from Mexico and then compare it to previously published data.

Maximum likelihood is a popular way of estimating branch lengths from complete mtDNA genome data. The method works by the program searching for the tree with the highest likelihood of giving the correct tree from the hypothesis that is inputted in the program. An advantage this method has is that it uses all the sites in a sequence in order to compute the most likely tree, instead of just the sites which contain different base pairs between sequences.

The problem with the above mentioned methods as described by Richards *et al.* (1996) is that these methods were developed for interspecific data and not for sequences which are closely related to each other. This problem was shown in trying to prove the out of Africa theory, where different authors such as Hedges *et al.* (1992) showed that there were a lot more equally good alternative trees which did not show the African origin of humans, even though later studies proved the out of Africa theory correct.

Analysis methods using networks instead of trees such as those developed by Bandelt *et al.* (1995, 1999) are more appropriate for mtDNA control-region sequences and Y-chromosome STR data as they can give an insight on recurrent mutations in the dataset. Using networks

meant that the original out of Africa theory was upheld and improved upon (Watson *et al.* 1997).

The use of networks is one of the best methods of inferring phylogenetic relationships as they present the relationship between different taxa in cycles which can show the possibility of a particular taxon arising from different genetic pathways instead of trying to resolve such ties (Bandelt *et al.* 1999). The method works by creating minimum spanning trees in a network and then adding median vectors in order to represent possible extinct or unsampled sequences in order to make the network parsimonious (Bandelt *et al.* 1999). Thus, this multistate method together with the Reduced Median method is superior to the previous mentioned methods due to how they tackle homoplasy.

Reduced Median networks use binary data instead of multistate data. This is not a problem since the vast majority of mutations observed in DNA are transitions. Due to this binary nature, ambiguous base pairs and deletions are removed from the primary analysis and then added subsequently in order to minimize the number of additional nodes (Bandelt *et al.* 1995). In order to reduce the network to the most probable evolutionary pathways, the weight of a character and the frequency of haplotypes are used (Bandelt *et al.* 1995). The authors showed that the system worked, and in subsequent analysis different researchers used networks in order to analyse both mtDNA and Y SNPs and also STRs. An advantage of this method is the creation of ancestral nodes in order to explain certain observations in the data set.

Thus Richards *et al.* (1996) used reduced median networks in addition to a molecular clock in order to determine that it is extremely unlikely that Neanderthal mtDNA survives in

modern humans due to different divergence times. The paper also showed the major female ancestral lineages of European populations, also showing how some of the ancestral lineages observed can be traced to Turkey and the Middle East. The conclusion reached through the use of Reduced Median Networks, and the molecular clock is that the majority of Europeans are descendent from people who settled in Europe in the Upper Palaeolithic. Migrations did occur from the East later, these migrations coincide with the spread of agriculture, but the influence on the modern population is not as previously thought (Richards *et al.* 1996).

The results from Richards *et al.* (1996) were confirmed in a later study (Richards *et al.* 2000) and improved upon, with the construction of a Western European skeleton network and the construction of Reduced Median Networks for each major European cluster. Such studies have shown the value of using networks in data analysis.

Whilst Principal Component Analysis (PCA) is useful when trying to observe clustering and relatedness patterns in multistate data, its limitations when compared to Networks is that it will not show how samples relate to each other and their possible common ancestral origin as it shows patterns at the level of populations, another disadvantage is that the method does not allow for the use of a molecular clock and thus the ability to date the emergence of different clusters from their ancestral origins. Both Networks and PCA are ways to explore data and try to observe patterns in the data and attempt to make inferences using these patterns

1.11: Mitochondrial DNA influences in Europe from the Near East

The most extensive work on mtDNA haplogroups in the Near East was conducted by Richards *et al*. (2000). In this study the authors used 1,234 samples from the Near East in

combination with European and Northern Caucasus samples in order to try and trace Palaeolithic and Neolithic European founder lineages in the Near East, from where it is thought the spread of modern humans in the Palaeolithic and then the Neolithic revolution progressed.

The results showed that there are more haplogroup sub-clusters in the Near East than in Europe and the Caucasus, suggesting that the Near East was the origin of the major European haplogroups. The age estimates of the haplogroups in Europe were also younger than in the Near East and the majority of haplogroups showed that they migrated into Europe from the Near East in the Lower Upper Palaeolithic. The results also showed that the rate of gene flow from Sub-Saharan Africa to the Near East was around 5% which Richards *et al.* (2000) comment is less than the gene flow that researchers estimated the Near East experienced from Europe which stands at between 10 to 20%. Haplogroups V and U5 also appeared to have originated in Europe according to the study.

The Neolithic influence in various regions was observed in the analysis by Richards *et al*. (2000) in that the Mediterranean region had a lower Neolithic component that that of central Europe, thus the Neolithic expansions affected genetically more the northern regions and the south Eastern regions of Europe than the Mediterranean and isolated areas such as Scandinavia and the Basques.

Turkey has always been an interesting country in terms of population genetics due to the different invasions and settlers it has had over time and the fact that it acts as a transition point between the West and the East from where such migrations would have passed.

Richards *et al.* (2000) mention that there is strong ground to assume that some back-migration happened from Europe back into Turkey

The study on 75 samples from Anatolia conducted by Mergen *et al.* (2004) was done in order to try and see what mtDNA could tell us about the region which connects Asia, Europe and the Near East. The authors suggest that the predominance of haplogroup H, which is the most common haplogroup in Europe, supports the claim by Richards *et al.* (2000) that this haplogroup originated in the Near East. The study also suggests that haplogroup U, which is a haplogroup found both in Europe and in Africa, spread into Europe from Northern Anatolia. A haplogroup frequency difference was also noted between Western and Eastern Anatolia. The authors explain that this could be due to historical migration from the Balkans into Western Anatolia, which the Eastern part of the region did not suffer as extensively from. A neighbour joining tree also showed that the region is most closely related to Central Asia than to Europe. This, the researchers explain, is probably due to the Mongol migrations from Central Asia which the region experienced. The analysis using neighbour joining trees as has been explained in the previous chapter should not be relied on as proof of relatedness.

The effects of isolation and population mixture in the same country in the Near East were seen when Gonzales *et al.* (2008) conducted a study in order to see the mtDNA haplogroups present in a sample of 145 people from the Jordanian population, and what relationship if any there was between these and neighbouring countries. Forty four of these samples were collected from a rural context whilst the rest were collected from an urban one. The samples collected from the urban context did not show significant difference from the populations of other nearby countries whilst according to the authors, the samples taken from the rural context showed genetic isolation. The authors say that this could be due to the physical

location of where the samples were collected from (the Jordan Valley), which is difficult to access and thus might have impeded migration.

In their study of Near Eastern mtDNA, Richards *et al.* (2000) found a high amount of gene flow occurring between Greece and the neighbouring areas. The authors attribute this to the historically known population movements between Greece and neighbouring populations including also back migrations from Europe to the Near East. Greece is also geographically positioned as a transient point for any migrations which have occurred from the Near East into Europe during the Palaeolithic and the Neolithic.

Kouvatsi *et al.* (2001) conducted a study which showed that the Greek samples were very homogeneous without any haplotypes being specific to a particular region of the country. When compared to other European and Near Eastern populations, the results showed that there was no genetic difference between the Greek population and their neighbours Bulgaria and Turkey, the homogeneity was even observed when compared to even far away European populations such as Tuscans, French and British populations. The problem with this study is the low sample size of the study at 54 samples, which is not representative enough of a country of almost 11 million people, and thus any inferences made from this study might not be sustained with a larger sample size.

An important study about the Greek population was the one conducted by Irwin *et al.* (2008) where the researchers sequenced HVR1 and 2 of 319 unrelated individuals from Northern Greece and 91 Greek Cypriots. The study concluded that whilst the Greek samples showed no significant difference when compared with Eastern Mediterranean populations and other

Greek samples which had been sequenced in other studies, these populations differed slightly from the Greek Cypriot sample.

Migrations which might have shaped the gene pool of modern day regions did not only occur during Prehistory, but historic migrations are also well documented. This means that Greece might also have influenced other areas in the Mediterranean where Greek colonies are historically and archaeologically known to have existed, such as Sicily and the South of Italy.

Sicily is interesting as it is possible that the island might still show some genetic traits left by previous colonisers such as the Greeks and the Phoenicians. The study of mtDNA was the pathway taken by Vona *et al.* (2001) in order to see the genetic composition of the Sicilian population, and try and determine its relationship with other populations, ranging from Spain to Turkey. The results from this analysis showed that there was a single instance of a huge population expansion in the island's history. The expansion increased the population from 300-900 individuals to 12,000 to 35,000 individuals, and occurred between 20,732 and 59,691 years ago (Vona *et al.* 2001). The authors also remark that there was no single nucleotide polymorphism (SNP) in the Sicilian study which had not been typed previously, and thus could be termed Sicilian.

In the study MtDNA SNPs showed differing degrees of variation when compared to other populations. Thus, the frequency of a mutation having occurred at mtDNA position number 16,126 in the Sicilian population was found to be midway between the frequency of this mutation occurring in the Eastern and western populations that were studied. The mutation frequency at position 16,311 was similar to that found in the Sardinian population, and also

to the Berbers and Turks (Vona *et al.* 2001). On the other hand the frequency of change at position 16,223 places it at the same level as that of the western European populations it was compared to.

The method used by Vona *et al.* (2001) is not the way relationships between populations should be studied. Specific mutations hold little value in determining how populations relate to each other when one does not look at the whole haplotype and how it forms part of a haplogroup. The study of specific haplotypes which might be found in both populations that are being analysed and the construction of networks are a much more useful tool in determining any link between different populations.

The Greek influence might also be seen in Southern Italy and might differentiate it from Northern Italy which was not historically colonised by the same populations. In their study on mtDNA variations in Southern Italy Ottoni *et al.* (2009) compared mtDNA data from the southern regions of Basilicata, Calabria and Sicily to other Italian and European published data. The study concluded that the haplogroup variability of these three regions was typical of the variability found in Western Europe. Haplogroups not of European origin were also found. These haplogroups were mostly of Near Eastern and North African origin and the authors ascribed them to the geographical position of Southern Italy at the cross-roads of the Mediterranean Sea. The southern regions also showed some affinity with Eastern Mediterranean regions with a high affinity in Basilicata. This affinity with Eastern regions also reflects the archaeological record of contact between the south of Italy and the Aegean and even the colonisation of Southern Italy by Greek city states.

As one moves further North from South Italy to North Italy, the genetic story changes as does the archaeology and history of the region. Thus in a study with forensic laboratory quality control data that was conducted by Turchi *et al.* (2008) in which the authors presented a mtDNA control region database of 395 samples which was collected from eight labs from central and northern Italy. It was concluded that the majority of the samples from Northern and Central Italy had West Eurasian mtDNA haplogroups, with haplogroup H being the most common, as with the rest of Europe and few North African influences were observed. The study also showed that haplogroup T was the second most frequent haplogroup.

The oldest database of mtDNA in Italy is the study by Di Rienzo *et al*. (1991) which used data from hypervariable region 1 (HVR1) of the mtDNA of a sample of Sardinian individuals as part of their database in order to study "branching patterns in the evolutionary tree for human mtDNA" this study nowadays can be said to be outdated due to the advances in methodologies in studying the specific regions, and also due to the limited length of the HVR1 region which was studied, which means that certain haplogroups are untraceable due to their mutation motifs not being in the range of the sequenced data.

1.12: Possible refugia and areas of isolation in Europe

Richards *et al.* (2000) point out that in the Basque samples, unusual frequencies of rare European haplogroups are an indication of genetic drift and a lack of Neolithic settlement when compared to the rest of the Mediterranean. The northern part of Spain has always been seen as a possible refugium of people during the last ice age, and thus Maca-Meyer *et al.* (2003) studied the Y chromosome and mtDNA features of an isolated northern Spanish

population called Pasiegos in order to see if any traces of such an event could be found. The Y chromosome results showed that whilst the population was isolated, this isolation had occurred recently. The study also showed that a considerable North African genetic input had occurred in the population which the authors attributed as probably being the result of the Moorish occupation of Cantabria. The authors also consider the high incidence of pre-V and V mtDNA haplotypes as an indication that Europe after the last ice age was recolonised from the South West (the Iberian Peninsula) with the Pasiegos and the Basques being the populations which started this expansion.

Côrte-Real *et al.* (1996) analysed 271 individuals from the Iberian Peninsula. This study showed the differences that exist between the Basque population and other populations in the region and also in the rest of Europe. The researchers also claim that the study shows a common point of origin of the Iberian population which occurred in the Upper Palaeolithic. The study also showed that the samples from Catalonia were genetically different from both other Iberian regions, and also from other European populations which were studied. The Andalucía samples on the other hand were highly divergent within themselves, showing all of the major European haplogroups within the group. Whilst this study is very useful, the fact that only 302bp were analysed limits the accuracy of the study, due to certain haplogroups not being identified as their motifs are situated outside the analysed region. The fifteen year old analysis also means that modern day haplogroup classifications were not used in the study.

On the other hand Flores *et al.* (2004) when looking at the Y chromosome haplogroups found that there is hardly any difference between the Basques and other isolated regions and the rest of the Iberian Peninsula and the difference that exists between different regions are

correlated with the distance between the regions. This according to the authors could be explained by either gene flow eliminating any differences which might have been present in the past, or from the differences in language and culture which are seen today having arisen after the Y chromosome genetic pattern of the Iberian Peninsula had already been established. This study is important due to the large population sample being studied at 568 individuals. This difference between the Y chromosomal and mtDNA studies might show a difference between males and females in this population were males migrate and thus dilute the genetic signal, whilst females stay in the region thus maintaining the signal.

Achilli *et al.* (2004) also looked at the possibility of a Franco-Cantabrian glacial refuge having influenced the European gene pool using mtDNA, the results were concluded from 62 samples of haplogroup H or closely related sequences. The results show that H1 and H3 closely follow the geographic distribution of haplogroup V and that these three haplogroups show an expansion of hunter gatherers after the last glacial period which repopulated Europe. Haplogroups H1 and H3 also have high frequencies in Sardinia, leading to the possibility that even Sardinia was a glacial refuge. This study whilst useful also suffers from low sample numbers when compared to other studies of the same region such as that conducted by Flores *et al.* (2004).

Thus, Sardinia as mentioned in the study by Achilli *et al.* (2004) has also been considered as a possible refuge or a place of genetic isolation, and the island has a history of isolation when compared to other Mediterranean islands, especially the mountainous interior of the island which different colonisers have always found difficult to subdue. This is also reflected in the Sicilian dialects which are far removed from modern day Italian. Calò *et al.* (2008) give a genetic review of the Sardinian population in which they show that classical

gene markers show internal heterogeneity in Sardinia which is due to isolation in some areas, but extensive mixture with other populations from around the Mediterranean in others. Even autosomal alleles show Sardinia as being different from other European populations, due to genetic drift, isolation and the genetic heritage left by people who dominated the island through time. The isolation is best defined in Nuoro, where the population is different to both other Sardinian regions and also to foreign ones.

The authors of the above study also discuss the reasons why mtDNA studies show Sardinia as having European haplogroups, but the frequencies of some of these such as haplogroups U and J are only half of what is found in the rest of Europe. The frequency of haplogroups range from region to region in Sardinia, thus showing genetic isolates where some of the haplogroup frequencies are very low, and populations with extensive contact with the mainland and other regions where the same haplogroup frequencies are high.

A study conducted by Fraumene *et al.* (2003) on the mtDNA gene pool of the Sardinian population, with an emphasis on Ogliastra, showed that Ogliastra was genetically homogenous. Ghirotto *et al.* (2010) tried to determine whether the Sardinian population in two different samples, one from Gallura and the other from Ogliastra, can be linked to the inhabitants of the islands during the time of the Nuraghe in the Bronze Age, using various computational models with both modern and ancient mtDNA. The results showed that the population of Ogliastra might be a descendent of Nuragic Sardinia, whilst the people from the other region of Gallura were more 'Europeanised' by gene flow coming from Italy. This study is interesting as it tries to look at mtDNA through different historical phases, which is more accurate than trying to infer the past by looking at modern day DNA alone. The

problem with such a study is that the researcher must always try and ensure that the aDNA is truly aDNA and not modern day contamination.

1.13: African mtDNA in the Mediterranean region

The North African populations as expected do not show the same haplogroup frequencies that European populations do, but the influences of Sub-Saharan Africa and historical migrations such as the Arab migration into the region have left their influences which are still visible in the genetic record.

A recent study that was that done by Fadhlaoui-Zid *et al.* (2011) used mtDNA in order to determine the population structure of North Africa. The results of this study showed that there is a divide between the Libyan and the Egyptian populations which typifies the haplogroup divide seen between the Eastern and Western North African region, with different haplogroups being prevalent in different regions. The Libyan sample in this study showed a low haplotype diversity of only 61%. The study also showed a prevalence of different haplogroups between East and West North Africa which can be associated to different phases of the known history of Africa, although this does not mean that the haplogroups have necessarily emerged in these phases, as prehistory could have influenced the African landscape and the lack of dating methodologies might not show it. This study is important as it is the first study to include a Libyan sample amongst the populations, and the high number of samples used in this study (around 2,300 samples) gives a better resolution to the study than some other contemporary Mediterranean studies do.

Turchi *et al.* (2009) analysed 120 Tunisian and Moroccan persons in order to determine their mitochondrial haplogroups. The results showed that haplogroup L was found in almost half of the Tunisian samples and in 25% of the Moroccan samples. This is consistent with the two nation's geographical position just above sub-Saharan Africa, and thus sub-Saharan haplotypes were expected. The authors also comment that the majority of haplogroups were still of Eurasian origin. This study suffers from low sample sizes with 64 making up the Tunisian samples and 56 the Moroccan one. Another problem with this study is that the samples are not from the whole regions of the countries but only North Tunisia and northern-central Morocco. This means that the results of the study though valid cannot be applied to all of the regions of the country since the non-sampled regions might show different patterns, especially the south of the countries which might reveal closer links to Sub-Saharan Africa due to their closer geographic position to the region.

1.14: The influence of North Africa on European populations

The close geographic proximity of certain European regions such as the Iberian Peninsula and the southern part of mainland Italy and the island of Sicily to the North African coastline and also some shared history between the regions has led to the probability of admixture between the populations. This admixture has been looked for using mtDNA and the Y chromosome as genetic markers in order to attempt to discern the influence North Africa had on the male and female lineages of certain populations.

The Côrte-Real *et al.* (1996) study on the Iberian Peninsula demonstrated that a small amount of the haplotypes found in the Iberian populations had shared lineages with North African haplotypes, but these were few in comparison to the ones showing shared lineage

with the rest of Europe. The authors thus conclude that the Moorish occupation of the Iberian Peninsula left little trace genetically in the modern population. As mentioned earlier this study whilst valid suffers from the fact that only 300bp of HVR1 was studied, and also from the study being dated and thus the nomenclature used in the study has changed over time.

The Portuguese population is interesting, since although not strictly Mediterranean the country shares a lot of its history and prehistory with Western Mediterranean countries, especially its neighbour Spain. Pereira *et al*. (2000) studied the distribution of mtDNA lineages of the Portuguese population. The Portuguese population displayed a higher diversity than the Spanish population; this is also accentuated by the higher number of lineages of African descent which are found in Portugal at a higher frequency than in other European populations including Spain. The diversity in lineage frequencies between north, central and southern Portugal was not statistically significant. The authors concluded by saying that whilst Portugal shared the mtDNA diversity of Europeans, the presence of haplogroup L lineages is synonymous with the African slave trade whilst the North African U6 lineage can be traced to the Moorish invasion, even though this haplogroup is found exclusively in the northern part of the country. Whilst this study is a very good attempt at explaining the origins if the L and U6 lineages, one must note that these lineages might have entered Portugal earlier than in the periods given by the authors since no dating of lineages was conducted. Whilst the explanation given by the authors is the most plausible, other migrations can never be discounted.

Plaza *et al*. (2003) also conducted a similar study were the matrilineal relationship between Western North Africa and the Western European populations was explored. The results

showed that the Western Mediterranean European populations share the same haplogroups as the rest of Europe with the addition of the L haplogroups which are found at low frequencies in the Western Mediterranean European populations. This contrasts with the Western African populations were the presence of L lineages shows that about a quarter of the gene pool has a Sub-Saharan origin (Plaza *et al.* 2003). Another observed difference between the two regions was that whilst the African populations were heterogeneous when the AMOVA was conducted, the European populations under the same analysis were shown to be more homogeneous. The study is very well conducted but as mentioned by the authors themselves; analysis of complete mitochondrial genomes might alter the findings of the study (Plaza *et al.* 2003).

The analysis of Vona *et al.* (2001) using mtDNA HVR1 showed that the Iberian peninsula was not the only European region that shows admixture with North Africa, but that Sicily possesses traits from both Europe and North Africa. This is shown due to the fact that haplotypes not normally present in Italy and other European populations, but which are present in the Berber population, can also be found in Sicily. This shows that Sicily was affected genetically by the populations which invaded it and settled in for a long time, such as during the Arab period. An interesting fact is that some unique haplotypes found in the Sicilian samples can only be found elsewhere in Sardinia. The analysis according to the authors also confirms the historical position of Sicily as being midway between the Middle Eastern and the European populations (Vona *et al.* 2001). Unfortunately this study suffers from a low sample size of 49 samples and from the fact that these were all sampled from an area near Palermo, thus these results might not be representative of the whole Sicilian island, but the inferences made in the study can only be applied to the particular village (Alia) situated near Palermo.

A mixture of both autosomal STR markers and mtDNA analysis was used by Romano *et al.* (2003) in order to try and shed light on the population genetics of Sicily. The authors say that due to their higher resolution, the markers they are using should give more accurate results than other markers which had been previously used to study the population of Sicily. An interesting part of the project is that surname data had also been collected and analysed.

The samples the authors obtained were divided into 6 geographical regions, covering the whole island, and each sample came from a person whose grandparents had also lived in the area (Romano *et al.* 2003). Tests on 9 STR polymorphisms and mtDNA haplogroup frequencies showed when analysed that Sicily is genetically heterogeneous. The authors attribute this heterogeneity to the complex history of Sicily involving migration and genetic drift, rather than to selective effects. Romano *et al*. (2003) also observed in their analysis that autosomal STR variabilities were higher than those of mtDNA, this they concluded may show that males migrated at a higher rate than females did throughout history. The results of the analysis also show that there was a higher genetic variability between the Eastern and Western samples than between the Northern and Southern ones. The analysis also included the divergence time of different samples, and from these results the authors say that there is an indication that the genetic composition of one of the towns (Sciacca) is mainly derived from settlements after the Roman conquest of Sicily, whilst all the other samples showed that they happened during the first and second century BC, thus after the Carthaginians settled in the west and the Greeks on the east of the island (Romano *et al.* 2003). Whilst this study is very useful in gaining an insight into population genetics in Sicily and is one of the most rigorously done study to date, the use of RFLP alone with no sequencing in order to

study the mtDNA data is nowadays antiquated as it does not give you individual mutations in HVR1 and HVR2 which allows the classification of haplogroups at a higher resolution.

1.15: Y-STR Studies in the Mediterranean

Capelli *et al.* (2005) analysed the variations that occur in Y chromosome lineages in the Mediterranean basin using 6 Y linked STRs and 16 SNPs, whilst also using previously published data. Principal components analysis led them to divide the Mediterranean region into four genetic groups. These divisions were described as being identifiable as a North African, Arab, Central-Eastern and finally a Western Mediterranean grouping (Capelli *et al.* 2005).

This study is also one of the few population genetics studies ever done which incorporated a sample of the Maltese population in its analysis. From the haplotypes that were determined from the Maltese sample, Capelli *et al.* (2005) concluded that the Maltese population clusters together with the Central-East Mediterranean group from a Y chromosome point of view (Capelli *et al.* 2005). The analysis also determined that the Maltese population was genetically nearer to North West Sicily, South West Sicily and Sardinia than to any other population which was included in the study. Y chromosome haplogroup frequencies on the Maltese island also revealed that a large range of haplogroups were present in the population, with the most common being haplogroup R1.

In the previous study 21.1% of the Maltese samples were defined as Y haplogroup J2, (Capelli *et al.* 2005). The authors associate these haplogroups with Arab and Arabised populations, with them also showing some presence in North African groups.

A point of interest that arises in the study conducted by Capelli *et al.* (2005) is that whilst North African populations have Y chromosomal haplogroup E3b frequencies of more than 50% and in some cases in the Mediterranean the marker can be found to be as high as 29%, in the Maltese sample the haplotype was found to be present in only 8.9% of the population (Capelli *et al*. 2005). This figure is much lower than would be expected due to the close proximity of the Maltese islands with its North African Mediterranean neighbours. The figure is even less than the estimation that Capelli *et al*. (2005) give of the male North African gene flow that occurred in the northern shores of the Mediterranean basin which they calculate to be between 15 and 27.1%.

Capelli *et al.* (2005) interpret the trends seen in the Western Mediterranean region, obtained from their study, as the result of gene flow due to western migrations by the Greeks and Phoenicians, and also the Arab conquest of the southern Mediterranean. Whilst this might well be the case, there is no way to differentiate without a molecular clock between these and other migrations which have occurred in history or prehistory, and this should have been mentioned in the analysis. At the same time, a higher number of Y-STR and SNP would have yielded a higher resolution and patterns would have been clearer.

Zalloua *et al.* (2008) also used Maltese samples when they tried to find signs of the Phoenician colonisations in different parts of the Mediterranean by using Y chromosomal haplotyping and STR analysis. In this study the authors identified a series of STR haplotypes which they defined as indicative of Phoenician colonisation, and measured the propensity of these STR signals in places which had been under Phoenician rule or had trade contact with them. The presence of the Y-chromosome haplogroup J2 seemed to indicate that this was a

marker with which one could trace the Phoenician expansion along the Mediterranean. These so called "Phoenician markers" were also found in the Maltese samples. This, according to an interview that Dr Spencer Wells gave, indicates that a lot of the modern Maltese population is a result of the Phoenician and subsequent Carthaginian colonisation of the island (http://ngm.nationalgeographic.com/ngm/0410/feature2/online_extra.html). Whilst this might be so, I am sceptical about the extent of the modern day Maltese population which is as a result of the Phoenician expansion. This is due to what is termed the 'Arab period' in Malta. In this episode it is known that the people who invaded Malta mainly came from Sicily and Tunisia (Luttrell 2002) both of which could have provided the Maltese with their "Phoenician markers" at that time instead of from the original Phoenicians as the Zalloua *et al.* (2008) claim, and there would be no way to differentiate between the two if both came through Tunisia without attempting to date clusters. This study is also plagued with inaccuracies such as the grouping of STR haplotypes together in order to make up for the one step mutation that the researchers expect to have happened from the time of the Phoenicians to the modern day (Zalloua *et al.* 2008). Another inaccuracy as mentioned previously is that without any dating methodology the relationships cannot be attributed to the Phoenicians but might have happened at any point in history or prehistory. The same pattern that is seen in what the researchers term as 'Phoenician' traits can also be seen in the Greek samples, which also adds to the difficulty in interpreting the results. The study on Malta will test whether or not the Lebanese population is closer to the Maltese population than the Tunisian population.

In an attempt to gain an insight into Greek population genetics from the male perspective, samples from Northern Greece had also been typed for 15 Y-STRs (Kovatsi *et al.* 2009). When the results were compared to Romanian samples and samples coming from Sardinia

and Ravenna, the analysis showed that whilst there was no statistical significance between the Greek and the Romanian sample, there was a difference between the Greek and the Italian samples.

Di Gaetano *et al.* (2009) conducted a similar study using Y chromosomal haplogroup and STR analysis on the Sicilian population. They concluded that Greece had a big impact on the Sicilian male population, contributing about 37% of the modern Sicilian Y chromosomal heritage. North Africa also contributed to the modern Sicilian population to a lesser extent according to the authors, who estimated the input to be of 6%. In their analysis the authors include a sample from Malta which was taken from Zalloua *et al.* (2008). The Maltese dataset groups closest to Western Sicily which is interesting when one considers that both regions were under the Phoenician and Carthaginian domination.

This result is interesting when one compares it to the Y-STR data from Kovatsi *et al.* (2009) which showed significant difference in the Y chromosomal data between the Greek and the Italian populations, thus Sicily which has a history of Greek colonization seems to still show traits of these occurrences whilst the Italian mainland has less visible links with Greece.

Robino *et al.* (2006) typed 255 Sicilian males for 8 Y-STRs and reported the haplotypes which emerged from the study. The results showed that there are statistically significant differences between areas from where samples were taken, this the authors note reflects the genetic heterogeneity of the Sicilian population which had been observed in previous studies using different markers. The results in this study are useful as it is the largest Sicilian male sample available to which the Maltese population can be compared.

On the opposite side of the Mediterranean differences were also noted between different regions in the Tunisian population. The Tunisian population was studied by Cherni *et al.* (2005) where the researchers analysed 11 STRs of four different groups from Tunisia. The results showed that whilst there was no difference between the Tunis, 'Tunisian Andalusian' and the Arab groups, these showed a difference from the other group (Berber) which was analysed. This difference, on the other hand, could not be seen between Berbers coming from Tunisia and those coming from Morocco, thus indicating that the two populations are homogenous, as are the three other populations in Tunisia. A criticism that can be made to this study is that whilst the total number of samples at 247 is amongst the highest for population studies in the region, the disparity in sample size between the different groups is great. Thus whilst the 'Andalusian' samples are made up of 132 individuals, the samples from the Berbers (30) and from the Arabs (31) are low in terms of size, and certain haplotypes present in the 'Andalusian' , Arab and Tunis samples might have not been sampled in the Berber one, thus showing a divide which in reality might not be there.

The results of the Arab 'Andalusian' samples in this study, which labels a group who are historically thought to have come to Tunisia after the Arab expulsion from Spain show that they are indistinguishable from their neighbouring populations. Bosch *et al.* (2001) explored the relationship that the Iberian peninsula had with its Southern neighbours, and the results of the study showed that there is a sharp contrast between the North African populations and the Iberian one genetically, with only about 10% of Y chromosomes in Iberia having been obtained from North Africa, and only about 7% of the Y chromosomes in North Africa having been obtained from the Iberian peninsula. This means that the samples labelled 'Andalusian' in the study by Cherni *et al.* (2005) had either been truly in Spain but did not mix with the local population, as such a mixture would have separated them from the other

North African groups in the analysis, or they were never in Spain and acquired that label from a legend which then turned into a tradition.

Part of the population of Libya was tested for 17 Y-STR markers (Elmrghni *et al.* 2012). The results were reported but were not compared to other populations from around the Mediterranean. It was the 238 males from Benghazi which were tested in this study that were used as a Libyan comparison to the present Maltese population study in order to see if any similarity between the two regions exists. Unfortunately population studies on the Libyan population are in their infancy and thus there are not a lot of sources to which the Maltese population can be compared.

Italy is an important place when trying to determine inter-regional differences in the same country since it has a varied history and differences in history and culture exist between the North, the South, and also amongst different regions which geographically might not be very distant from each other. Presciuttini *et al.* (2001) typed 8 autosomal STRs in 1176 samples from 10 Italian regions spanning the whole extent of the country. The study showed a low degree of heterogeneity between regions which the authors explained may be due to low sample numbers when compared to the whole of the Italian population. The created database is useful in order to make comparisons with other populations from the Mediterranean region such as the Maltese. If the low heterogeneity was not due to sample size, it means that the historical differences between North and South Italy are not reflected genetically.

A study with a larger amount of STRs analysed and concerning the Y-chromosome was the one conducted by Turrina *et al.* (2006). In this study 155 samples from North-Eastern Italy were typed for 17 Y chromosomal STRs. Unfortunately the results of this study were only

compared to a sample from the Austrian population. The results of the comparison with the Austrian samples typed for the same STRs showed that no significant difference was found between the populations. The authors also specify the importance of typing for a high number of STRs as this increases the resolution in a study.

Y chromosomal STR analysis on the Sardinian population, which is considered as a possible genetic isolate, was conducted by both Scozzari *et al.* (2001) and Ghiani and Vona (2002). The results showed that there is genetic heterogeneity in the Sardinian male samples between regions. This, when also compared to the Y chromosomal haplogroup analysis, shows that the Sardinian male population is composed of typical European haplogroups, even though in certain regions the ratios of these haplogroups are not typically European. The allele frequencies from Ghiani and Vona (2002) were used as a comparison population for the current Maltese population study

1.16: The use of alternative markers in population genetic studies of regions neighbouring Malta

Other markers have also been used in an attempt to gain a glimpse at the relationships of other regions in the Mediterranean. These studies are normally older than the mtDNA and the Y chromosome ones since the markers used are not considered to be as informative in recounting the story of how different populations interacted with one another and within the same population.

A very interesting study was conducted by Guglielmino *et al*. (1991) in which they tried to find out if by using surnames as genetic markers one can discover genetic patterns in a population, and if then these patterns might be attributed to a certain point in history. The

authors believe that surnames can be used as genetic markers since they are passed patrilinealy like the non-recombining parts of the Y chromosome. The study was conducted using the registers of consanguineous marriages (isonomy) between 1910 and 1964, thus the authors believe that since the marriages were consanguineous, which implies that they had lived in Sicily for more than one generation, the results give an indication about the genetic population of Sicily at the start of the 20th century (Guglielmino *et al.* 1991). The results showed that there are four different genetic regions in Sicily, with the largest separation being between the Eastern part and the Western part of the island. This division was found by the analysis of surnames. Gugliemino *et al*. (1991) noted that the areas in which Greek colonies had been set up corresponded to clusters formed with the Greek surnames used in the analysis, and that although this might be a coincidence, the authors point out that there is no reason why people with these surnames could not have migrated to other parts of the island during history, thus annulling the formation of the cluster (Guglielmino *et al.* 1991).

The use of isonomy in order to study populations is becoming more popular. In their review, Colantonio *et al.* (2003) showed how such studies have been used in order to show genetic trends or abrupt changes in the population. The authors also remark that studies have shown that surnames can be used in order to show geographic origins, as in the case of Switzerland where the surnames form clusters according to their place of origin (Colantonio *et al.* 2003). Thus the authors also mention the study conducted by Sykes and Irven (2000) which found that in an analysis of persons surnamed Sykes 43% of the individuals had the same MSY haplotype.

The analysis of surnames is very useful when it is used as an aid to genetics in order to reveal cultural elements of a population which might have an effect on the gene pool which hard science cannot reveal.

An attempt to shed light on whether there is this visible east-west genetic division in Sicily was attempted by Rickards *et al.* (1998). They did this by testing 2354 persons for 24 markers in order to see if such a division exists, and if it does, if that division could be attributed to the historic colonisations by the Greeks and Phoenicians of the opposing parts of the island. Their findings show that there is no discrimination in all of the analysed data between East and West Sicily, with the samples not forming a statistically significant difference between east and west (Rickards *et al.* 1998). The authors emphasis that analysis of their results shows that Sicily is one panmictic unit (there are no mating restrictions between the people of the island). Rickards *et al*. (1998) also say that the results they obtained are further strengthened by the fact that there is no obvious physical feature in Sicilian topography that might act as a barrier between parts of the population and their abilities to meet each other and thus mate (Rickards *et al*. 1998). In their conclusion the authors mention that DNA analysis of ancient populations from Sicily and other parts of the Mediterranean will mean that their analysis will be either confirmed or rejected.

An analysis of the data from previous studies that had been conducted in the whole of the North African region was conducted by Bosch *et al.* (1997). The different datasets were compared both against themselves, and also against other regions, in particular Europe. The markers that were analysed in the Bosch *et al.* (1997) study were what are termed classical markers which included blood groups, red blood cell enzymes, proteins and HLA antigens. The analysis, using neighbour-joining trees, showed that the Libyans and Egyptians

clustered very closely to European Mediterranean populations and Saudi Arabians, whilst the North West Arabs, Berbers and Tuareg clustered separately, and were distinctive from the Libyans and Egyptians. The authors conclude that the result could be due to migrations during the Neolithic that replaced the western populations of the region, whilst this did not happen in the east due to the terrain that allowed hunter-gatherers to be able to continue living without being replaced by farmers, or they themselves having to resort to agriculture. The authors also conclude that the Arab invasions that occurred during the seventh century AD, replacing the indigenous populations living there, could also be an explanation for the difference seen between Eastern and Western North African populations. It is hard to understand how the authors came to the conclusion of replacement of hunter gatherers by farmers from the patterns shown above.

The authors say that the more genes that are included in the analysis, the more Western Arab speakers appear close to Egyptians and Libyans, a fact that could be seen as evidence that genetic transfer did occur after the invasion between the then indigenous people and the Arabs; this transfer can also be pointed at culturally, since they adopted the language of the new arrival. Another thing that comes out of this study is that Sub-Saharan Africa plays a relatively small part genetically in terms of contributions to North Africa. This can be expected due to the Sahara desert acting as a barrier between the regions, and also since there is no record of any wave of migrations occurring from the south to the north. Turchi *et al.* (2009) have shown this last conclusion not to be true in their study of Tunisian and Moroccan populations using mtDNA with a Sub-Saharan influence being visible in the region.

This study also showed that Libyans are more closely linked genetically to Andalusians than more western countries like Morocco and Algeria are (Bosch *et al.* 1997). This is a very interesting point when one considers that the invasions into Spain occurred from the latter. Thus the authors conclude that the Arab invasion only had a limited impact on the Iberian gene pool. This study supports the findings of Bosch *et al.* (2001) which found when analysing the Y-chromosome that the Islamic period in Spain left only a minor contribution to the gene pool and that North African populations and Iberian populations originated from different expansions and migrations, even though they are geographically close to each other.

Another later study concerning North West Africa was also conducted by Bosch *et al.* (2000). In this study the authors used STR markers in order to get an insight into the genetic structure of the Maghreb region and how people who speak different languages relate to each other, and also how they compare to other populations especially in Iberia and Sub-Saharan Africa: this study had a higher resolution than the previous study conducted by the same authors. The results of the STR analysis showed that the level of heterozygosity along the populations was very similar except for the Mozabite population (Bosch *et al.* 2000). The level of heterozygosity of the samples was also similar to that found in Iberia and other regions. The authors ascribe the lower level of heterozygosity found in the Mozabites as due to genetic drift, since they are a very isolated population. The study by Bosch *et al.* (2000) that there is a separation between the North West African populations and the European region, and the closest European regions to that of North West Africa are those from the Iberian Peninsula. This result supports the findings of mtDNA and Y-STR studies that show that although there was a genetic contribution by North Africa to the genetic makeup of the

Iberian Peninsula, this was not in an extended way, and the respective populations are still rather distinct genetically.

The authors conclude that even though the different populations that inhabit the Maghreb have got different cultural traits, they are still genetically too similar to be able to be distinguished clearly from one another, thus no genetic differences could be implied between Arabs and non-Arabs in the region (Bosch *et al.* 2000). The authors also conclude from this that the Arabic invasion of the North West African region was more of a cultural colonisation rather than a genetic replacement, as seems to have happened in Eastern North Africa. This contrasts with the results of Fadhlaoui-Zid *et al.* (2011) which shows that a divide exists. The difference in conclusions might be due to Bosch *et al.* (2000) using autosomal markers which means that genetic traits get diluted due to both the mother and father contributing to them. Thus the resolution is not as great as when one looks at just the matrilineal or the patrilineal line since a mixture of both means that the chances of replication slippage is coming from both sides since both the mother and the father are contributing STRs . Thus the higher power of discrimination of autosomal STRs to uniquely identify an individual which make them useful in court cases, act as a hindrance in long term generation studies such as this when a specific set of STRs are looked at together to determine how they relate to other sets of STRs.

An attempt to study the Maltese population was conducted by Cassar *et al*. (2008). In this study the authors analysed the allele frequencies of fourteen autosomal STR loci in 157 unrelated Maltese individuals. The frequency levels of five of the STRs were then compared to those found in other populations from around the Mediterranean region which had previously been published. The reason given for using only five of the STRs typed was that

these are the most common ones found in publications, and thus are comparable to the data of a wide range of publications. The results obtained from this study were similar to those of Capelli *et al*.(2005), as the Maltese population was deemed closest to the Sicilian population, with the authors commenting that the two North African populations included in the study were genetically distant to the Maltese one.

Cassar *et al*. (2008) comment that the results of their study also showed that the Italian population had closer ties to Greece than to the Maltese and Sicilian populations. The authors add that the Maltese population is genetically heterogeneous and remark that it expanded exponentially by means of founder effects to the present day size from a smaller population. Historical records also support this idea, where it is seen that the population of Malta starts to increase at a faster rate after the arrival of the Knights of Saint John on the island, when pirate attacks became less common (Abela 1647). It was also during this period that the northern part of the island started to be settled, as before this it was uninhabited due to the population preferring to live close to the three castles on the islands where they could take refuge during a pirate attack (Abela 1647).

Cassar *et al*. (2008) also mention the constant contact the Maltese had with Sicilian mariners during the time of the Knights; the authors say that their analysis is also supported by the number of Sicilian surnames present on the island. Whilst the use of only 5 STRs to compare them to other populations is understandable due to the lack of published data, this also means that the results do not possess the higher resolution that a greater number of STRs might have yielded, and thus patterns which might have been visible in this analysis might dissipate at a higher resolution.

1.17: Aims of the study

The aims of this study can be divided into various questions:

1. The most immediate aim of this study is to determine if aDNA survives in the Maltese climate, and whether aDNA can be extracted from Late Roman and Prehistoric sites in Malta, this research question needs to be answered since no aDNA studies have ever been conducted on samples coming from the Maltese archaeological record.

2. The study also aims at answering the research question: Where males and females buried together in the sites under study? Sex identification through genetic means would help the researchers of the sites in question with their sex identification based on anatomical attributes.

3. Another research question which will be attempted to be answered in this study is: Were people related to each other buried in the same complex? And if they were where they placed next to each other in the sites under study? This research question is important as kinship relationships in burials cannot be identified through traditional archaeological research methods, and thus aDNA might give us a glimpse into ancient Maltese burial practices which traditional archaeology would not be able to do.

4. Another research question which needs answering is: Were there any genetic differences in who inhabited the urban area and the rural area in Malta during the Roman period? This would give a more general picture of the Maltese islands in the

Roman period, and whether there was a difference between the people who lived in the city and the ones who lived in the rural areas as can be seen through documents later on in Maltese history during the Norman period where the ruling elite lived in the fortified city whilst the indigenous population lived in the countryside.

5. The most pressing research question that is to be answered is: To which other populations is the Maltese population closest to genetically? An answer to this will be attempted from a matrilineal viewpoint by comparing the mtDNA HVR1 of the Maltese population with other populations from around the Mediterranean region. This way a profile of the female line of the Maltese population can be acquired, whether the haplogroups found are typically European or from elsewhere, and which haplotypes are found in other populations from around the Mediterranean will be determined. This is important due to traditional historical records concentrating mostly on the male perspective, with events concerning the females of a population normally included more as an afterthought.

6. The same thing will be done with the male lineage where the Y chromosome STR data of the Maltese population will be compared to similar data from other populations in the Mediterranean region. This analysis will also see if the conclusions of previous studies on the male Maltese population are replicated.

7. A research question which will also be investigated is: Are there any traits found in the Maltese aDNA samples which can be also be found in the modern Maltese population? This will be done by comparing the results from the aDNA samples with those from the modern DNA of Maltese samples. Any traits which carry on from the Maltese aDNA samples into the modern ones would be a sign of genetic continuity in the population.

8. The last research question which will be attempted is: Are any of the population bottlenecks which can be discerned from the history of the Maltese islands visible genetically? Due to the history of the Maltese islands, where several times the islands suffered major population loss, bottlenecks would be expected and would confirm these events genetically. Thus a comparison of the genetic history and the written history of the island would be made in order to see if there are any similarities or differences.

Chapter 2: Materials and Methods

2.1: Acquiring permission to sample the inhumations from the Maltese burial sites

The Superintendence of Cultural Heritage, which is the Maltese governing authority with regards to matters of national heritage, was contacted by the researcher in early September 2007 in order to discuss the possibility of using inhumations found in an archaeological context in Malta as a pilot study in order to understand the extent of the survivability of aDNA in burials on the Maltese islands.

A meeting was held by the researcher with the Superintendence of Cultural Heritage, in order to discuss with the Superintendence the sites which could be made available in the study and during this meeting the initial two sites of Tal-Gardina and Tal-Barrani, which were investigated in this study, were proposed by the Superintendence of Cultural heritage on the basis that they were both roughly of the same age but that they had been excavated over 15 years apart. These sites proved ideal since they were useful for the research questions that the researcher wanted to tackle in the project and also they had the additional benefit that they came from the same time period and were excavated years apart, thus raising more research questions that could be tackled in the study. During the meeting an arrangement was made with the Superintendence to write a proposal specifically for the two sites. The proposal was written and this was accepted by the Superintendence after they were given specific guarantees about technical issues such as the storage of samples after the DNA analysis, and that they would have access to any results that were obtained from the study involving Maltese human remains.

During April the following year, the Tal-Gardina site was visited and a meeting with the osteologist working on the site was also held, in order to ensure that the best samples were chosen for the study and also to make sure that the samples were excavated and held in the appropriate manner in order to ensure the best preservation of aDNA and to limit the possibility of modern DNA contamination from the excavators. A meeting was also held with the principal investigator who had conducted the excavations at Tal-Barrani in order learn more about the site.

The Superintendence of Cultural Heritage also asked the researcher to prepare a national plan for aDNA analysis in Malta, including regulations as to how this should be conducted, and what minimum standards the laboratories that are used in such an analysis should have. This national plan was prepared and submitted to the Superintendence (Appendix 1).

The Kercem samples were the last samples to be collected after being offered by the Superintendence as a means of providing a benchmark for the earliest periods in Maltese prehistory from which aDNA results could be obtained. Three inhumations were sampled; each inhumation had a long bone sampled and a molar. The reason for this was that the dental studies on this site had been finished, whilst it had not for the previous two sites. Of all the sites, the Kercem samples were the easiest to obtain due to the Superintendence's willingness to push back the boundary of aDNA studies in Malta after they were informed of the Tal-Gardina and the Tal-Barrani preliminary results. The researcher when accepting to conduct the research on Kercem explained that the retrieval of aDNA from this site would be difficult due to its age and it having been flooded during the excavation.

2.2: Obtaining the modern Maltese samples

Direct sampling of people from the Maltese population was considered at the start of the project, but in order to avoid the possibility of pathogens from the saliva swabs or the blood samples, Malta Biobank, which is based at the Department of Genetics of the University of Malta, was contacted in order to see if it was possible to obtain the samples from their bank. This also avoided ethical issues such as consent forms (which had been collected by the mentioned department) and whilst keeping all donors anonymous ensured a transparent way of disposing of samples after the project. After an explanation of the project, the University of Malta kindly agreed to provide the researcher with the samples on the agreement that they will receive a copy of any publications which might come out from the project. The researcher also gave a talk as part of their Life Sciences seminar series. The DNA bank provided the researcher with 150 randomly selected anonymous Maltese male DNA samples of varying concentration that had been obtained from blood and were stored in Tris/Borate/EDTA (TBE) buffer solution. The selection of only male samples was made in order to obtain both mitochondrial and Y chromosomal DNA from the samples. These were the samples used in the subsequent modern Maltese population mitochondrial and Y chromosomal DNA analysis. The last condition that was made for the samples was that after the project concluded, any remaining samples would be returned to the Malta Biobank.

2.3: Obtaining the archaeologists' samples

The archaeologists and osteologists who excavated the sites at Tal-Gardina and Kercem were contacted by the Superintendence and were asked to submit a DNA sample in order to compare their sequences with any results that might come out from the study in order to try and eliminate the possibility of modern contaminants being mistakenly identified as aDNA.

In the case of Tal-Barrani this proved to be more difficult due to the time which had elapsed between the excavation and the DNA study of the site, but whilst not all of the excavators were reachable in order to submit a samples for the DNA analysis, a DNA sample was obtained from the principal investigators and those who came mostly in contact with the inhumations in order to try and reduce the chances of contamination. Not being able to test people who have come in contact with samples is an inherent problem in ancient human DNA studies involving excavations which have been conducted decades before the aDNA analysis.

The DNA testing process involved sending each person to be tested two cheek swab testing kits (Master Amp^{TM}) together with a consent form which also gave instructions on how the samples should be taken. The pair of cheek swabs for each individual was needed in order to ensure that if no DNA was successfully extracted from one of the cheek swabs, the other cheek swab would yield results. The cheek swabs together with the consent forms were then given back to the researcher so that the mtDNA haplotyping could commence. After the conclusion of the project the DNA samples and cheek swabs were destroyed as per the agreement on the consent form.

2.4: Sample conditions

The Tal-Barrani samples were the most physically robust. The bones were thick indicating that the people had seen manual labour during their time, and the bones were still quite robust and were the best preserved. The smell of collagen whilst sawing off the bone to take the samples was a good indicator that biomolecular material was still present in them and that they might yield aDNA. The bones from Tal-Gardina were the second most well

preserved. Whilst the bones were not as thick as the Tal-Barrani samples and some of them had been broken post mortem, they were still in a good state of preservation. The physical difference between the Tal-Gardina and the Tal-Barrani samples was that the Tal-Gardina bones were not as thick as the latter ones. Keeping in mind that the site is just outside the old city wall, one arrives at the conclusion that this was due to the persons buried there not being involved in regular manual labour like the Tal-Barrani ones. The Kercem bones were the least well preserved amongst the three sites. This could have been due to their age and also due to the fact that the site had been waterlogged for long periods of time. This did not bode well for the aDNA studies. Another factor which was noted was that whilst collagen degradation could be smelled to varying degrees in the other two sites whilst sampling the bones, the Kercem samples did not give that distinctive smell whilst they were being sampled.

2.5: Precautions and quality control

When working with ancient human DNA, the quality control of the work is of utmost importance since contamination is the most common factor for obtaining false positive results. Precautions were taken throughout the whole process of the analysis in order to avoid contamination issues.

A guide to how archaeologists should excavate human remains that are destined for aDNA investigation has been published by Brown and Brown (1992). This was handed over to the osteologists and excavators who were working on the Tal-Gardina, Tal-Barrani and Kercem specimens in order for them to be able to preserve the samples as well as possible for aDNA work. Whilst some of the things written in the guide, such as the wearing of gloves whilst excavating, were adopted, unfortunately other things, such as the wearing of face masks

during the excavation were deemed inappropriate by the Superintendence of Cultural Heritage for use in the Maltese climate and thus not adopted. The methodology used in excavating and analysing specimens is always an issue between the archaeologists and the aDNA researchers, as some of the things that are done during and after excavation, such as the removal of soil from bones by brushing with distilled water, is not conducive to good aDNA research. In the case of Tal-Barrani, since the excavation had been conducted a long time before the study, only the person studying the bone material before it was passed on to the researcher could follow the guide.

The sampling of the bone material was conducted in a room; where everybody who was in close proximity to the sampling was wearing a full body suit, hair net, facemask, eye protection, foot protection and was also double gloved. This was done in order to ensure the minimum possibility of contaminating during the sampling process.

The tools used for the sampling were also used exclusively for this purpose. This ensured that no contamination from the tools would occur. The hacksaw and all blades used were wiped vigorously with 'DNA Away ™' (Molecular Bioproducts) between each sampling from the same skeleton and left to air dry. This product has been shown to remove roughly two thirds of surface-attached DNA when wiped softly (Champlot *et al.* 2010), and the author of the study (Champlot *et al.* 2010) maintains that a more vigorous and extensive wiping (as was performed on the Maltese samples) should be more efficient. The wiping with DNA Away™ and the subsequent irradiation with UV in the laboratory (which in the study conducted by Champlot *et al.* (2010) prevented the amplification of 95.6% of the DNA), was done in order to eliminate or reduce as much as possible the chance of contamination. This in addition to the removal of the outer 2mm of the bone should have

reduced the chance of contamination from the same skeleton as much as if a new hacksaw blade was used, since the new hacksaw blade would have still not been guaranteed to be DNA free and the place where the samples were taken from in Malta was also not DNA free. After each session of sampling from the same skeleton the hacksaw blade was also discarded as an added security before a new skeleton was sampled.

2.6: Precautions in the laboratories

The aDNA laboratories where all the work was conducted were only used exclusively for this purpose. Thus the extraction room where the samples were crushed and DNA was extracted from them was physically separate from the main laboratory were PCR amplifications were conducted. A separate room existed where the master mixes for aDNA amplifications were prepared and then added to the extract before being brought up in the main laboratory for PCR amplification. This room was not even on the same floor as the main laboratory in order to reduce the chance of cross-contamination as much as possible. Another precaution was that in the main laboratory the PCRs were always conducted in thermocyclers that were exclusively used for aDNA analysis. Whilst the thermocyclers themselves will amplify DNA, not using them for modern DNA analysis meant that the copy numbers the PCR started with was always lower and the DNA more degraded than if it had also been used for modern DNA. This strategy might not necessarily be effective in reducing contamination, but was standard practice in the laboratory.

Both the master mix preparation room and the extraction rooms were constantly irradiated by UV when not in use, thus ensuring that any DNA that was present would be cross-linked and thus not amplifiable. This UV irradiation came from a strategically placed tube in both rooms which was only turned off when people were about to enter in order to start working

inside the rooms. The work surfaces of the laboratories were also wiped with a 30% concentration of bleach before the start of every day, in order to remove any contamination. All apparatus in the rooms and the cabinets where the work was conducted were also wiped with 'DNA Away™' (Molecular Bioproducts) as a precaution. Everything that was used during work in the aDNA laboratories such as pipettes, tips, tubes and pens was irradiated with UV by means of a Stratalinker (Stratagene) for 10 minutes on each side before being used.

The two rooms also had a positive pressure system in place in order to prevent any dust found in the corridors from entering into the rooms. Another feature used to reduce the chances of contamination was the antechambers of the two aDNA rooms. The antechambers served a dual purpose; they served as a safety area separating the UV irradiated rooms from the corridor, and they were also a place where the suiting up of the personnel working in the rooms was conducted.

The floor of the room where the extraction work was conducted was washed at the end of each day when bone crushing was conducted, and once every week additionally. The floor of the room where the PCR was set up was washed once a week. This was done by the researchers themselves in order to make sure that only personnel who were trained in the use of the rooms had access to them.

Extract and PCR blanks were done for every step of aDNA analysis. This ensured that any contamination as a result of a mistake in the execution of the procedures was immediately found out without going through the whole process before the contamination was observed, which is what happened when individual samples were contaminated instead of any of the

master mixes. A positive control was never used alongside the aDNA PCRs. Whilst this meant that if no DNA was obtained from a reaction one was in doubt whether it was PCR failure or no DNA being present in the sample, the introduction of a positive control in the same area where no DNA should be present would have increased the risk of cross-contamination. It would not then have been possible to claim that no modern DNA had been introduced in the aDNA areas of the building and that no work involving modern DNA was conducted in these rooms.

The cloning of every aDNA amplicon was also conducted in order to ensure that any result was not a mixture of more than one source of DNA, because from analysis of the clones different DNA sequences in the same sample could be observed. This also helped in differentiating between original aDNA and modern contamination.

The utmost effort was made to obtain a cheek swab of every person who came in contact with the bones and teeth from which the samples originated. The DNA extracted from the swabs was then amplified. Thus a profile of these persons, who could potentially be sources of contamination, was taken (Appendix 2). This precaution meant that all the sequences of amplified DNA could also be checked against this information in order to ensure that any modern contamination from a person who had worked with the samples was immediately identified. Unfortunately due to the age of some of the samples, not all the people who were involved with the excavation could be tested. This is a limitation in aDNA studies which is often encountered when dealing with samples which have been excavated a long time before the DNA studies are conducted.

2.7: Ethidium bromide precautions

Precautions were also taken in order to ensure that the health and safety standards expected in the laboratory were maintained. Gloves and safety glasses were used at all times whilst working in the laboratory. The gloves were also changed every time ethidium bromide was handled during the preparation of the agarose gels, when this substance was still in use in the laboratory. The reason is that ethidium bromide is considered mutagenic. A pipette was also exclusively used for pipetting ethidium bromide into the melted agarose solution.

When visualising the gels using a transilluminator (Geneflash) a glove had to be removed by the operator, and the exterior surfaces of the transilluminator were handled with this hand, whilst the hand still having the glove was used in order to position the gel in the transilluminator and then remove it at the end of the visualisation. This procedure was followed in order to make sure that the pictures taken of the agarose gels which contained the bands of DNA were not contaminated by ethidium bromide since these would eventually be taken out of the laboratory. The transilluminator had also a safety feature that switched off the UV source automatically when the door was opened. The writing up area was also situated outside the laboratory; this reduced the time spent in the lab whilst at the same time ensured that persons were not exposed to unnecessary risks and kept the chances of contamination to a minimum.

2.8: Sampling

The bones and teeth deemed suitable for DNA analysis were chosen together with the osteologist based on their appearance, the conditions that they were found in, and their

general robustness. Long bones were sampled preferentially to other parts of the skeleton due to their ease of cutting, thickness when compared to other bones, and also durability.

The researcher cut the fragment of bone for each sample used in this analysis by means of a hacksaw or an electric drill. The hacksaw was used when the researcher was allowed to take large samples from the skeletons, whilst the electric drill was used in order to make a small square shaped incision in the bone and remove a fragment from this window when the researcher was not allowed to take a large fragment from the skeleton. The use of the electric drill also gave an indication of the biomolecular content of the bone, and if the smell of collagen was detected during the drilling it was taken as a good sign that the bone would probably yield aDNA.

Bones which already had a breakage in them were sampled preferentially to other parts of the same skeleton. Whilst this might have increased the risk of contamination, they were chosen in order to limit the physical damage to the remains as far as possible. Each bone to be sampled was laid out on a clean surface on top of aluminium foil, and then the hacksaw was used to cut an approximately two-inch fragment from the bone. After every cut, the hacksaw blade was wiped vigorously with 'DNA Away^{TM}' (Molecular Bioproducts), and the gloves were changed in order to avoid any cross-contamination that might occur between different samples of the same skeleton. At the end of each skeleton sampled the hacksaw blade was changed and the researchers changed their gloves again. The researchers also attempted to keep the bone dust that arose from the cutting to a minimum, and the piece of foil on which the sample was placed whilst being cut was also changed after every sample. The samples were clearly labelled and packed in DNA-free bags for their voyage to the

laboratory in Manchester. The samples were then held in a fridge at 5°C until they were ready to be processed.

All the teeth which were sampled in this study were all still attached to the jawbone; these teeth were chosen in order for them to have a greater degree of protection from contaminants and also to ensure that they had not been handled previously. The teeth were carefully dislodged from the jawbone and bagged. Whilst having the tooth still attached to the mandible does not ensure that no contamination is present, it reduces the chances of it. The problem with the sampling of teeth was the lower amount of biological material that can be extracted from them.

2.9: DNA extraction methods

2.9.1: Bone preparation

The outer 2mm surface of each sample was first scraped off using a scalpel. Each side of the bone fragment was then UV irradiated for ten minutes (254nm, 120,000μJ cm^{-2}) and then crushed by means of a mallet in a sealed UV irradiated DNA-free bag which had been inserted into another similar bag. Three 0.25gram powdered bone samples were weighed out from each sample and placed in 1.5ml centrifuge tubes to await DNA extraction. Any excess powder from the crushing of the bone was also retained for possible future work on the same samples. This process was conducted in a clean room wearing a full body suit that was worn exclusively during bone crushing, and not during the DNA extraction from the bone. All of the tools used during the procedure were UV irradiated for ten minutes on each side (254nm,

120,000μJ cm^{-2}). This was done in order to ensure that any modern DNA contamination that may have been present would be cross-linked, and thus not amplifiable.

2.9.2: Removal of tooth dentine

When it was possible DNA from a tooth of the subject was extracted independently to the extraction from a bone. The selected teeth were always still attached to the jawbone of the skeleton as a precaution against modern DNA or PCR inhibitors entering the tooth from the root cavities which are exposed. The advantage of sampling teeth is that the DNA from the tooth dentine can serve as an independent sample from the same subject. The disadvantage of teeth is that less powder is obtained from them on which to do the extractions, and thus the success rate of aDNA analysis using teeth is less than that of using long bones. The tooth was first UV irradiated for 10 minutes, and then the exterior was wiped with a 40% bleach solution. After this it was dipped in distilled water for 10 minutes without submerging the roots in order to remove any traces of bleach. The tooth was then cleaned by the application of a phosphoric acid gel (DeTrey©) to its exterior without touching the root cavities, this was done in order to clean the exterior surface from debris and destroy any DNA present there. After the gel was wiped away, the tooth was then submerged again in distilled water for 10 minutes in order to remove the phosphoric acid, and then air-dried. One of the root cavities of the tooth was then widened by the use of a dental drill bit (Dentsply, Maillefer). This was done by rotating the drill bit slowly by hand, and the powder which started coming out at first was discarded. As the drill bit started entering the tooth, the drill bits then started being changed into thinner ones in order to reach further inside. When a drill bit reached the pulp of the tooth, the powder produced by the drilling started then being collected in a clean microfuge tube which had been previously irradiated for 10 minutes under UV. Care was

constantly taken whilst hand drilling as the tooth would become increasingly weak as the drilling continued and there was a risk of the tooth breaking in half. The drill bits were discarded after the powder was obtained, and were not used for extracting powder from another tooth. One ml of a solution of 0.5M EDTA pH8.0, 0.5% SDS and 100µg Proteinase K was added to the powder and the tube was left in a shaking incubator at 55^{o}C for 22 hours. After this the tube was centrifuged at 13,000rpm for 3 minutes, and the resultant supernatant was taken and purified using a QIAquick gel extraction kit.

2.9.3: DNA extraction and purification from bone and teeth

A 1ml solution made up of 0.5M EDTA pH8.0, 0.5% SDS and 100µg Proteinase K was prepared for each sample. The solution was then UV irradiated (254nm, 120,000µJ cm^{-2}) for ten minutes in order to make sure that it did not contain any amplifiable DNA. As a further precaution against contamination, throughout the process, two extract blanks were prepared and were treated the same way as the samples. The solution was added to the pre-weighed crushed bone samples, which were then agitated continually at 55°C and 650rpm for 24 hours in order for the Proteinase K to digest the calcified matrix barrier in the osteocytes (Li 2009).

The samples were then centrifuged in a microfuge at 2000 rpm for 5 minutes, after which 2.5ml of buffer PB (Qiagen) was added to 0.5ml of the supernatant. These were then mixed and 0.70ml of the solution was pipetted into a Qiaquick gel extraction kit spin column and centrifuged at 13,000rpm for 1 minute in a microfuge. The solution that passed through was discarded and the process was repeated until all of the solution had been passed through the spin column. Then 0.75ml of buffer PE (Qiagen) was added to the spin column and the

flow-through was discarded after it was centrifuged at 13,000rpm in a microfuge for 1 minute. A further centrifugation under the same conditions was conducted in order to ensure that the entire buffer had been removed.

The spin column was then put in a clean 1.5ml microfuge tube, and 50µl of buffer EB (Qiagen) added. The spin column was left to stand for 1 minute, and then centrifuged in a microfuge at 13,000rpm for 1 minute. This process was repeated for a further time. Next, 22µl of the extract was transferred into a new centrifuge tube and taken for analysis, whilst the rest of the solution was kept in a freezer for future use. This was done in order to preserve the samples, since every time they are defrosted DNA damage occurs.

2.9.4: Modern DNA extraction using chelex

DNA was extracted from all of the persons available that came in contact with the samples during excavation and in the laboratory. For the persons in the laboratory this was done using the method described by Walsh *et al.* (1991) where the person's saliva was first collected in a microfuge tube and then was centrifuged for 10 minutes at 10,000rpm. After the removal of the supernatant the remaining cells were placed on ice and 500µl of 10% chelex bead suspension was added to them. The sample was then vortexed in order to make sure that the chelex beads would not collect at the bottom of the tube. The solution was then placed in boiling water for 10 minutes in order to break open the cell walls, after which it was placed on ice for a further 1 minute. The sample was then centrifuged in a microfuge at 13,000rpm for 30 seconds; the supernatant was taken and stored at -20^{o}C until analysis.

For the archaeologists who were excavating on site the same method was used, but instead of saliva samples the MasterAmp^{TM} buccal swab testing kit was used. This kit was used as it bound the DNA for a month at room temperature and six months at 5^{o}C without the DNA suffering degradation. This meant that the samples were stable enough from the time the swab was taken to the time they were posted from Malta to Manchester. The buccal swab brush was cut and placed in the chelex solution and vortexed and boiled according to the above protocol. Before the final centrifugation the brush was removed.

For the analysis on the modern Maltese population no extraction was needed, since the samples which were kindly provided by the genetics laboratory at the University of Malta had already been extracted using their in-house protocol and were handed in ready to be used.

2.10: Polymerase chain reactions

2.10.1: Ancient DNA markers used in the study

Eight multiplex amplifications were used in the analysis of the Maltese samples (Table 2). These multiplexes contained seven primer sets which covered the whole extent of mtDNA HVR1 and HVR1I. These were the primers from which the greatest numbers of sequences were expected, since they targeted mtDNA which is found at a much larger quantity than nuclear DNA.

Ten sets of primers in the multiplexes were used in order to target a variety of short tandem repeats and single nucleotide polymorphisms in autosomal DNA, whilst seven primer sets targeted regions in the Y chromosome. As mentioned previously, a primer set targeting the amelogenin gene was also used in an attempt to identify the sex of the burials at all the sites being studied.

These multiplexes had been designed by Dr Abigail Bouwman and used in a previous study concerning samples from Mycenae (Bouwman *et al.* 2008) in the same laboratory. This had the added advantage of the multiplexes having already been optimised and tested out on archaeological samples of the same age as the Maltese samples, thus the researcher knew that the multiplexes worked.

Multiplex name	**Primer pair name**	**Locus targeted by primer**	**Location of targeted sequences**
C1	MtF	16210-16340	MtDNA
	GD	d1s656	Autosomal DNA
C2	MtV	182-318	MtDNA
	YB	dYs389	Y-Chromosome
	YA	dYs393	Y-Chromosome
	GL	CD4	Autosomal DNA
D	MtA	15996-16114	MtDNA
	MtG	16028-16195	MtDNA
	MtD	16271-16420	MtDNA
	YZ	M35	Y-Chromosome

	GO	TH01	Autosomal DNA
	GN	d8s1179	Autosomal DNA
E1	MtC	16147-16294	MtDNA
	GA	Amelogenin gene	Y and X chromosomes
E2	MtW	16-132	MtDNA
	YH	dY460	Y-Chromosome
	GE	d10s2325	Autosomal DNA
	GM	d3s1358	Autosomal DNA
G	mtW	16-132	MtDNA
	YF	dYs391	Y-Chromosome
	YH	dY460	Y-Chromosome
	GE	d10s2325	Autosomal DNA
	GK	VWA	Autosomal DNA
	GM	d3s1358	Autosomal DNA
H	GF	d8s535	Autosomal DNA
	GJ	d6s366	Autosomal DNA
	GK	VWA	Autosomal DNA
	GL	CD4	Autosomal DNA
J	GG	d5s8181	Autosomal DNA
	YG	dYs426	Y-Chromosome
	YY	M173	Y-Chromosome

Table 2: Details of the multiplex PCRs

2.10.2: Amplification

The reaction mix for the PCR was made up in another clean room separate from the extraction room. All material used for this part of the procedure except the AmpliTaqGold® DNA Polymerase was irradiated with UV (conditions: 254nm, 120,000μJ cm^{-2}) for 10 minutes in order to ensure that no cross-contamination occurred from the previous time the reagents had been used. A 30% bleach solution was also used to clean the work surfaces. A list of the $MgCl_2$ concentrations and the annealing temperatures used for each multiplex can be found in Table 3. The DNA was then amplified using the multiplexes listed in Table 2.

Multiplex	**Annealing temperature (^oC)**	**$MgCl_2$ (mM)**
C1	55	2.0
C2	54	2.5
D	53	2.0
E1	55	2.0
E2	55	2.5
G	55	2.0
H	56	2.0
J	55	2.0

Table 3: $MgCl_2$ concentration and annealing temperature required for specific multiplexes.

The typical composition of a single PCR was a 50µl solution containing

1× AmpliTaqGold® DNA polymerase buffer

XmM $MgCl_2$: Where X is the multiplex $MgCl_2$ concentration listed in Table 3

0.2mM Each dNTP (dATP, dGTP, dCTP, dTTP)

0.1µg /µl BSA

2ng /µl Forward primer

2ng /µl Reverse primer

2.5µl Sample

0.05U /µl AmpliTaqGold® DNA Polymerase

The PCR cycles used were:

Program 1: 1 cycle at 94°C for 7 minutes

Program 2: 35 cycles of

X°C for 1 minute: Where X is the multiplex temperature listed in Table 3

72°C for 1 minute

94°C for 1 minute

Program 3: 1 cycle of

X°C for 2 minutes

72°C for 10 minutes

10°C for 1 minute.

2.10.3: Modern mtDNA markers

Two sets of primers were designed using the Primer3 program. One of these primer pairs covered HVR1 (Figure 5) whilst the other covered HVR2 (Figure 6). The primers were optimised with an emphasis on trying to remove primer dimer formation as much as possible, and then the samples were amplified using a *Taq* DNA polymerase which did not require heat activation before use. This was done due to cost considerations. The amplification was conducting using the following reaction conditions:

1× Standard *Taq* Buffer including Mg^{++} (New England Biolabs Inc.)

0.2mM Each dNTP

0.1µg /µl BSA

2ng /µl Forward primer

2ng /µl Reverse primer

2.5µl Sample

0.025U /µl *Taq* DNA polymerase

The reaction was then topped up to 25µl with H_2O

The Mg^{++} concentration in 1× Standard *Taq* Buffer is 1.5mM. The reaction was prepared on ice and then inserted in a pre-heated thermocycler at 95^{o}C and the following amplification conditions were used:

HVR1:

Program 1: 1 cycle of 94^{o}C for 5 minutes

Program 2: 25 cycles of

53^{o}C for 30 seconds

72^{o}C for 30 seconds

94^{o}C for 30 seconds

Program 3: 1 cycle of

53^{o}C for 2 minutes

72^{o}C for 7 minutes

15^{o}C for 1 minute

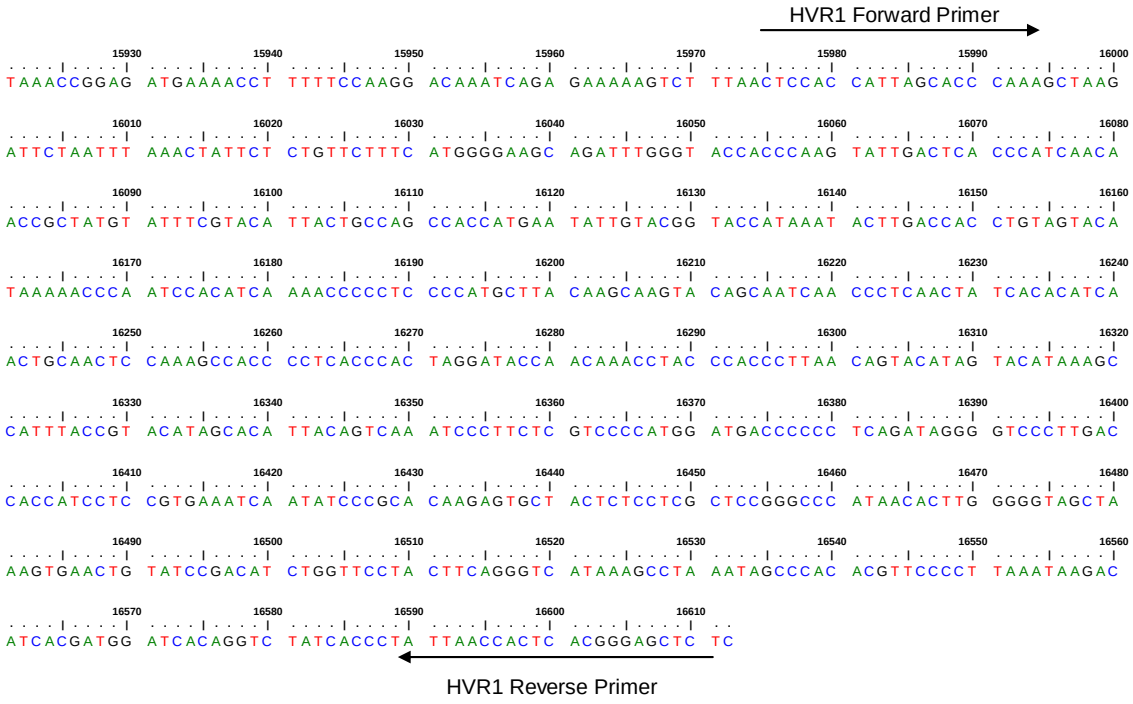

Figure 5: Region of HVR1 covered by the forward and reverse primers.

HVR2:

Program 1: 1 cycle of 94°C for 5 minutes

Program 2: 28 cycles of

60°C for 30 seconds

72°C for 30 seconds

94°C for 30 seconds

Program 3: 1 cycle of

60°C for 2 minutes

72°C for 7 minutes

15°C for 1 minute

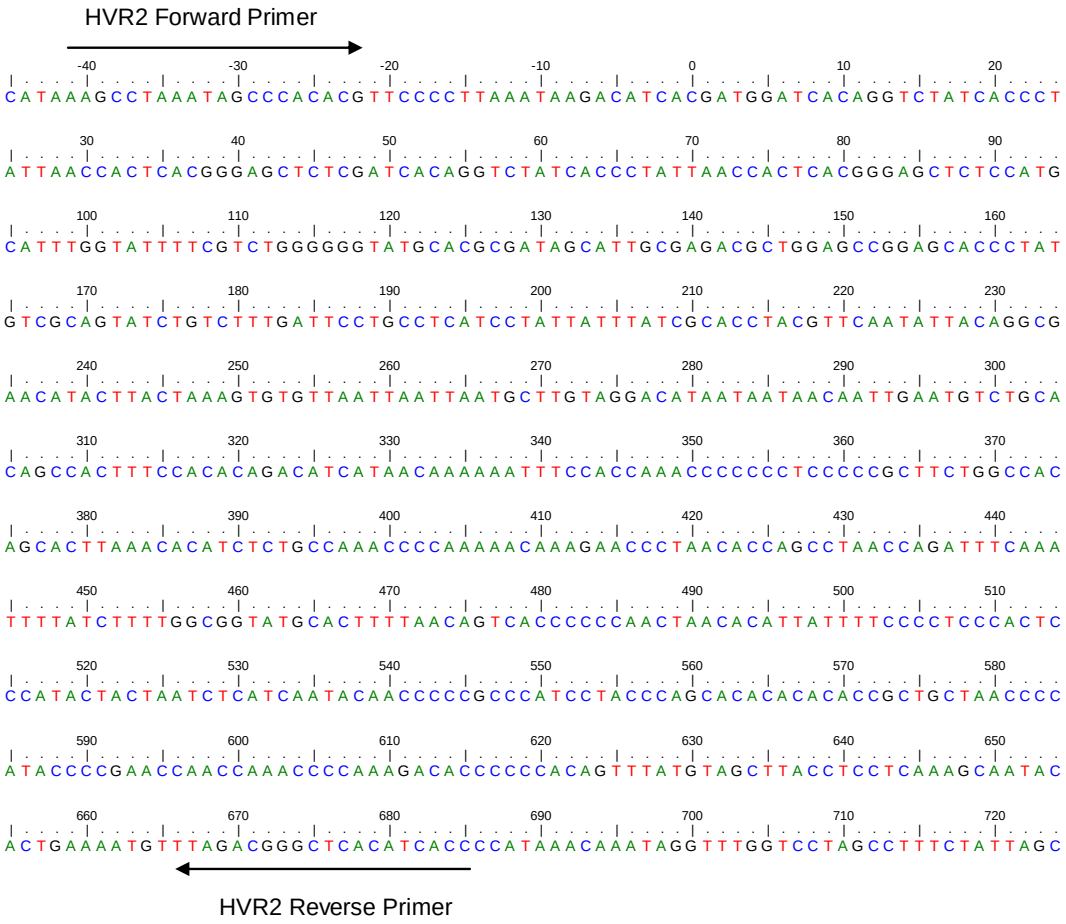

Figure 6: Region of HVR2 covered by the forward and reverse primers.

The results were then visualised in a 2% agarose gel (Figure7), and if a band of the correct size was present the reaction was sent off for sequencing (GATC Biotech using the Applied Biosystems 3730XL DNA Analyzer using the BigDye© Terminator sequencing standard kit 3700/3730/3730XL systems). The raw sequencing data can be found in Appendix 3

The HVR1 primers amplified a region of 637bp whilst the primers targeting HVR2 amplified a region of 685bp.

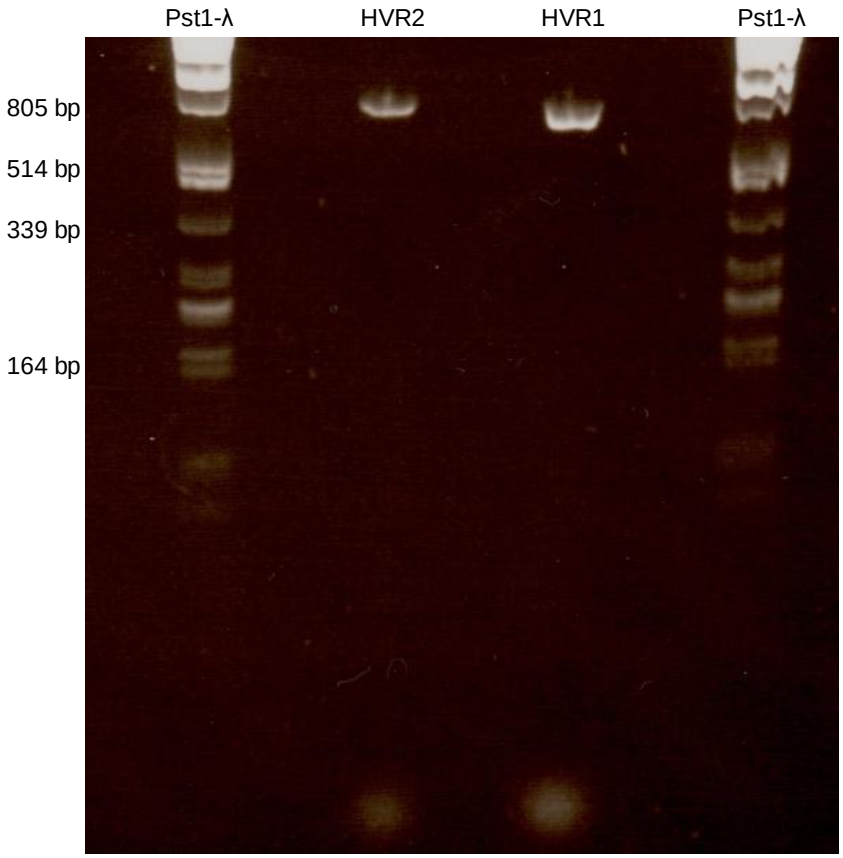

Figure 7: 2% Agarose gel showing both HVR1 and HVR2 amplifications with PstI-λ as size marker.

2.11: Preparation of DNA size markers

DNA size markers were used to determine the size of the product being visualised in the agarose and polyacrylamide gels. Size markers were prepared in a volume of 100µl by adding 20µg of λ DNA (Fermentas) to 1×*Pst*I buffer, adding 40U of *Pst*I and topping up with MilliQ water. The solution was then placed overnight in a 37^{o}C water bath in order to allow for the restriction digest to occur.

The Bioline HyperLadder^{TM} 100 bp ladder was also as a marker. The advantage of having a commercial marker was that it eliminated slight inconsistencies present in the markers prepared in the laboratory. A reference diagram showing the size and position of the bands when the markers were run on a 3% agarose gel was used as a reference.

2.12: Gel electrophoresis

The 2% agarose gels used to visualise DNA were prepared by weighing 1g of agarose powder in a flask and then adding 50ml of 1×TBE (90mM Tris (hydroxymethyl) aminomethane, 90mM boric acid and 2mM EDTA). This was then heated in a microwave oven on medium setting for about 2 minutes until the agarose dissolved. Then 2µl of ethidium bromide (10mg ml^{-1}) or 0.7µl of Gel Red was added; the liquid was then poured in a cassette and allowed to cool down, thus forming the gel. Fifteen µl of sample and 2µl of loading dye (30% glycerol, 0.25% bromophenol blue and 69.75% H_2O) were loaded in each well. These were run in the gel tank until the dye reached about an inch from the end of the cassette. In the extreme right hand well of every gel that was run, 15µl of *Pst*I-λ or

HyperLadder™ together with 2µl of loading dye was loaded. The ladders served as a marker from which the sizes of the bands under analysis were determined.

2.13: Comparison between routine reagents and new substitutes used in the laboratory

2.13.1: Comparison of electrophoresis buffers

A sodium borate buffer was used when the DNA was just being screened for its presence, size and position on a gel, but not when DNA was going to be extracted from the gel for cloning. The reason why DNA was not extracted from sodium borate gels was that the researcher could not find any mention in the literature of cloning being conducted from DNA extracted from sodium borate gels, and thus whether the DNA run through a gel with this buffer would be viable for cloning.

The advantage of sodium borate gels is that they can be run at a higher voltage than gels made from TBE. The reason is that the sodium borate gels do not absorb as much heat as TBE since TBE creates a runaway positive feedback loop whilst sodium borate buffer does not (Brody and Kern 2004). The reason given by the discoverers of this buffer is that TE and TBE buffers were the first ones used for protein separation and then when they were applied for DNA analysis, nobody had ever tried finding better buffers specifically for DNA screening (Brody and Kern 2004).

The lower heat absorption of sodium borate buffer meant that a 2% agarose gel electrophoresis could be conducted at a voltage of 15V cm^{-1} for sodium borate gels without

running the risk of it melting, when compared to the safe 9V/cm used for TBE. This meant that the time to conduct the screening was reduced from 90 minutes to 45 minutes. Sodium borate buffer was also used as the buffer when screening DNA using polyacrylamide gels, and here also it performed better than TBE due to its increased thermal stability. Lithium borate is a better buffer than sodium borate in terms of heat dispersal capacity (Brody and Kern 2004), but since the chemicals needed to make it are hazardous and it was expensive to purchase, the sodium borate buffer was used instead as the best alternative buffer.

The sodium borate buffer was made in the lab by adding boric acid to 10mM of sodium hydroxide until a pH of 8.0 was reached (Brody and Kern 2004).

2.13.2: Comparison between Gel Red and ethidium bromide

At the start of the project ethidium bromide (Invitrogen) was used in order to make the DNA fluoresce after binding when excited with UV radiation under a standard UV transilluminator. When Gel Red™ (Biotium) was introduced on the market it substituted ethidium bromide in its use in the lab. The advantage of Gel Red is that it is not mutagenic and thus safer to use for the operator when compared to ethidium bromide. This also impacts on its cost as Gel Red is cheaper to dispose of then Ethidium Bromide. According to manufacturer claims, which were also observed in the laboratory, Gel Red is also more sensitive than ethidium bromide although it has the same visible light excitation wavelength (Biotium 2009). Whilst this might have been construed as an advantage, practically, it impaired work in the lab, as bands of DNA which were not visible with ethidium bromide started becoming visible with gel red, but unfortunately when excised, purified and cloned invariably they gave no result due to the amount of DNA which had been visible with Gel

Red not being sufficient in order to conduct cloning. On the other hand this feature also served as an advantage when conducting aDNA analysis, as very faint bands in controls which would not have been visible with ethidium bromide were now seen and thus the amplifications were discarded. The higher sensitivity of Gel Red also meant that a very small amount was needed in order to stain the gels when compared to ethidium bromide (0.7µl/50ml of Gel Red instead of 2µl/50ml for ethidium bromide).

Agarose gels were pre-stained with Gel Red instead of post-stained, as this reduced the time of the experiment, and no significant difference in the quality of the gel between pre-staining and post-staining could be determined. Post-staining was conducted in the case of polyacrylamide gels as the manufacturer of Gel Red did not recommend pre-staining in this case, the same staining method was employed when ethidium bromide was used.

2.14: Extraction of DNA from agarose gels

When primer dimer was shown to be present in a PCR product, instead of direct purification from the reaction, the targeted DNA band was cut from an agarose gel using a sharp blade and then placed in a pre-weighed microfuge tube. The DNA was then extracted from the agarose slice using a QIAquick gel extraction kit (Qiagen) using the following protocol.

Three volumes of buffer QG were added to 1 volume of the gel slice (where 1µg was equal to 1µl). The mixture was then placed in a water bath at 50^{o}C and vortexed every 2 minutes until the agarose melted. One volume of isopropanol was then added to the mixture and the mixture was then placed in a QIAquick spin column and centrifuged at 13,000rpm for 60 seconds. The flow-through was discarded and 0.5ml of buffer QG was added to the spin

column and then the column was centrifuged again for another minute, with the resulting flow-through again being discarded.

A volume of 750µl of buffer PE was then added and the flow-through was discarded again, after which the spin columns were centrifuged again at 13,000rpm in order to ensure that any traces of ethanol which might have remained would have been eliminated. The spin column was then transferred to a clean centrifuge tube and 50µl of the elution buffer EB was added to it after which the spin column was centrifuged again, with the flow-through being now kept for further analysis.

2.15: Direct PCR purification

When primer dimer was not present, and only the band of interest was amplified, the PCR product was purified using the QIAquick PCR purification kit (Qiagen). Following the protocol that came with the kit, 75µl of buffer PBI was added to 15µl of the PCR product. The colour of the resulting mixture was observed, and whenever this turned orange, instead of the yellow of buffer PBI as it should have been, 10µl of 3M sodium acetate at pH 5.0 was added in order to change the colour back to yellow. Each sample was then placed in a QIAquick spin column. Each of the columns had their own 2ml collection tube, and the samples were centrifuged in a microfuge for 60 seconds at a speed of 13,000 rpm. After the flow-through was discarded, 750µl of buffer PE was added to the QIAquick spin column, and the samples were centrifuged for an additional 60 seconds. After the flow-through was again discarded, another microfuge centrifugation was done for 60 seconds at a speed of 13,000 rpm.

The QIAquick spin column was then placed in a clean 1.5ml microfuge tube, as any remaining buffer would have inhibited the reaction. 50µl of buffer EB was then added to the samples in order to elute the DNA.

The reason for using the QIAquick PCR purification kit is that with this method any DNA less than 40bp in the reaction was removed. This included any unused primers still in the reaction, their primer dimers, dNTPs and any other impurities which were less than 40bp in length.

2.16: Comparison between different kinds of silica gel spin columns

Two different brands of silica gel spin columns were tested in order to see the difference in how much DNA they retrieved when they purified the same DNA sample whilst using the same reagents. The reason for this test was that one of the brands cost a quarter of the price of the name brand. The results of five different samples processed in this way were compared, and the results showed that the name brand collected a mean of 31% more DNA then the generic brand. After this finding the researcher started to use the Qiagen spin columns when the samples which were being processed were aDNA and thus the maximised collection was needed, whilst the generic brand name was used when purifying modern DNA, due to the higher concentration of DNA that results from a modern PCR when compared to a PCR that had aDNA as a template.

The manufacturers of the EconoSpin^{TM} columns (Epoch Life Sciences) maintained that the reagents of the name brand could be used with their spin columns, whilst at the same time the manufacturer gave a protocol that could be used in order to make the reagents. All the

comparisons were conducted using the standard reagents included with the name brand spin columns, in order to make sure that what was being compared was the effectiveness of the silica gel spin columns themselves, not the different reagents of the two products.

2.17: Cloning of the PCR products

2.17.1: TOPO® cloning kit

The cloning kit that was used in this analysis was the TOPO® TA cloning kit for *Taq* polymerase-amplified PCR products (Invitrogen). The procedures listed below are very similar to the protocols found in the user manual and just variations of them. The protocol was not followed to the letter due to the specific product size needed, cost, and time constraints that were involved in this project.

The LBA medium (Luria-Bertani agar) was obtained from stores where it had been prepared by technicians by mixing 10g of tryptone, 5g yeast extract and 10g sodium chloride in 950ml of deionised water. To this mixture 15g of agar was added. This LBA was then placed in a microwave oven, and defrosted for 40 minutes in order for it to liquidise. The LBA was then left to cool, after which 400µl of 100µg/ml ampicillin was added to it. The LBA was then poured into 10cm plates and left to solidify. When it had solidified, the plates were inverted and incubated at 37^{o}C until these were ready for use.

The sample was then prepared for cloning by adding 0.25µl each of vector (pCR®2.1-TOPO®) and salt solution (1.2M NaCl, 0.06M $MgCl_2$), together with 1µl of the amplified

PCR in a microfuge tube, this was then left at room temperature for 30 minutes for the ligation reaction to occur. Whilst the protocol says that 5 minutes was enough for the reaction to occur, the reaction was left for 30 minutes due to the protocol also stating that the longer the reaction is allowed to continue at room temperature, the greater the yield obtained would be.

Ten microlitres of TOPO10F' (Invitrogen) competent cells were then added to each sample after being allowed to defrost on ice for 10 minutes. The cells needed to be defrosted for 10 minutes before being used as they were kept at the bottom of the -80^{o}C freezer in order to maintain their integrity. The mixture was left on ice for a further 30 minutes, so that the *Escherichia coli* cells would become acclimatised. The samples were then inserted in a water bath at 42^{o}C for 30 seconds and immediately afterwards put back on ice for 2 minutes. This was done since the resulting thermal shock induced the cells to take up the vector which by now contained the strand of amplified DNA. Seventy five µl of SOC medium (20g Bacto-tryptone, 5g Bacto-yeast extract, 0.5g NaCl, 2.5ml of 1M KCl, and H_2O for a volume of 1000ml) was then added to each tube, these were then placed in a shaking incubator at 37^{o}C for 1 hour.

During incubation period, solutions of 40mg/ml of X-gal in DMF and 100mM IPTG were prepared. Forty microliters of each of these mixtures were spread evenly around the plates. The plates were then incubated for a further 15 minutes at 37^{o}C.

The samples that had been placed in the shaking incubator were then spread onto the plates. A blank reaction was also prepared and plated in the same way as the others. This plate served as a control for the experiment, which ensured that the colonies that were growing

were the ones targeted, and no colonies were present due to contamination during cloning, thus in the control only blue colonies should be seen. The plates were than incubated overnight at 37^{o}C in order to allow enough time for the individual cells to form colonies.

The plates were then checked. Blue colonies were ignored, as this meant that they had taken up the vector but it did not contain an insert. From the white colonies visible on the plate the most isolated and largest 10 to 12 were selected. Sometimes if fewer white colonies were visible than ten, light blue colonies were also selected as there was a possibility that these had taken up the DNA. The colonies that were selected from each plate were then amplified using the master mix described below with the M13Forward and M13Reverse primers (Invitrogen).

The reactions were made up to 25µl volume comprising:

2.5µM 10X Taq buffer

1 unit *Taq* DNA polymerase

2µM Each dNTP

50ng M13Forward

50ng M13Reverse

The PCR cycles consisted of:

Program 1: 1 cycle of

94^{o}C for 10 minutes,

Program 2: 25 cycles of

94°C for 1 minute

55°C for 1 minute

72°C for 1 minute

Program 3: 1 cycle of

55°C for 2 minutes

72°C for 10 minutes

15°C for 1 minute

The lid of the thermocycler was always preheated to 105°C before the start of the run. This denatures the DNA and thus reduces random PCR amplifications.

In order to check if the cloning had been successful, 2% agarose gels were prepared and 5µl of each amplified sample was run on the gel. The samples that had not taken up the vector together with the DNA, or that had taken a fragment of DNA which was not of the expected size, and thus not useful in the project, were discarded.

The resulting PCRs were then purified by means of the QIAquick PCR purification kit (Qiagen) as described above. Fifteen microlitres of the purified sample was sent off for sequencing (GATC Biotech).

A sequence alignment editor (BioEdit) was used in order to analyse the sequences obtained. When the chromatogram was of poor quality, manual alignments were made since it was noticed that BioEdit is not efficient when the sequences are quite dissimilar to each other. Sometimes the chromatograms had to be discarded due to them being of a very low quality.

2.17.2: StrataClone™ PCR cloning kit

The Strataclone™ PCR cloning kit was used as the TOPO® cloning kit over time started to yield lower amounts of colonies and the ones which had yielded white/light blue colonies would sometimes not contain the insert but instead a mixture of primer dimer. The cloning protocol was modified in an attempt to yield the desired colonies but this did not have any effect. The new cloning protocol was very similar to the TOPO® one but although not perfect it yielded better results.

The LBA plates were prepared by adding 5ml of 10mg/ml ampicillin to each 500ml of LB medium.

The ligation reaction was as follows:

3µl StrataClone™ cloning buffer
2µl PCR product
1µl StrataClone™ vector mix.

The reaction was mixed together gently in the prescribed order and then left at room temperature for 5 minutes after which it was placed on ice. A tube (50µl) of the supplied competent cells was thawed by being left for 10 minutes on ice and then 1µl of the cloning mixture was added to the tube and mixed gently without any pipetting. The reaction was then incubated on ice for 20 minutes during which time the SOC medium, which is a bacterial growth medium, was warmed up to 42^{o}C.

When the 20 minutes had elapsed the reaction was heat shocked for 45 seconds at 42^{o}C, after which it was incubated on ice for 2 minutes. The heat shock treatment was needed in order for the cells to absorb the vector. A volume of 250µl of SOC medium was added to the mixture and the mixture was placed in a shaking incubator at 37^{o}C for 1 hour. Meanwhile solutions of 40mg/ml of X-gal in DMF and 100mM IPTG were prepared. Forty microliters of each of these mixtures were spread evenly around the LB-ampicillin plates. The plates were then incubated for a further 15 minutes at 37^{o}C. After the 1 hour in the shaking incubator, 100µl of the reaction was spread on each plate, and the plates were left in a 37^{o}C incubator overnight. Ten light blue or white colonies were then selected and amplified with the M13Forward and M13Reverse primers (Invitrogen) using the same reaction and under the same conditions as with the TOPO® cloning kit.

The amplifications were then screened in a 2% agarose gel and the ones which had taken up the insert were sent off for sequencing in 96 well plates. Whilst the amount of light blue and white colonies was greater than with the TOPO® cloning system, and more clones took up the vector with the insert, the results from the sequencing showed that a high percentage of primer dimer was still present in some of the samples.

2.18: Sex identification

The amelogenin gene is very useful in sex identification of archaeological samples as it is found on both the X and Y chromosomes. The difference between the two chromosomes is that the amelogenin gene on the X chromosome has a 6bp deletion when compared to the Y chromosome. When a result was obtained by means of the amelogenin gene, the identification which could be made was either male, or probably female. This is because

whilst males would show two distinct bands, females would only show one band, although the latter result meant that allelic dropout could not be excluded.

In order to determine the 6bp difference which distinguishes whether a sample is a probable female or a male the resulting band in the agarose gel was excised, purified and run in a polyacrylamide gel. The polyacrylamide gel was made up in a vertical electrophoresis system using the following recipe:

2.5ml of 10x TBE

6.25ml of 40% acrylamide/bis-acrylamide, 19:1 solution

50µl TEMED

260µl 10% AMPS

The sample was then loaded onto the gel together with *Pst*I-λ and a 100bp ladder to act as markers and was run at 15 V/cm for 60 minutes.

The gel was then post-stained using Gel Red after which it was visualised in a transilluminator (Figure 8).

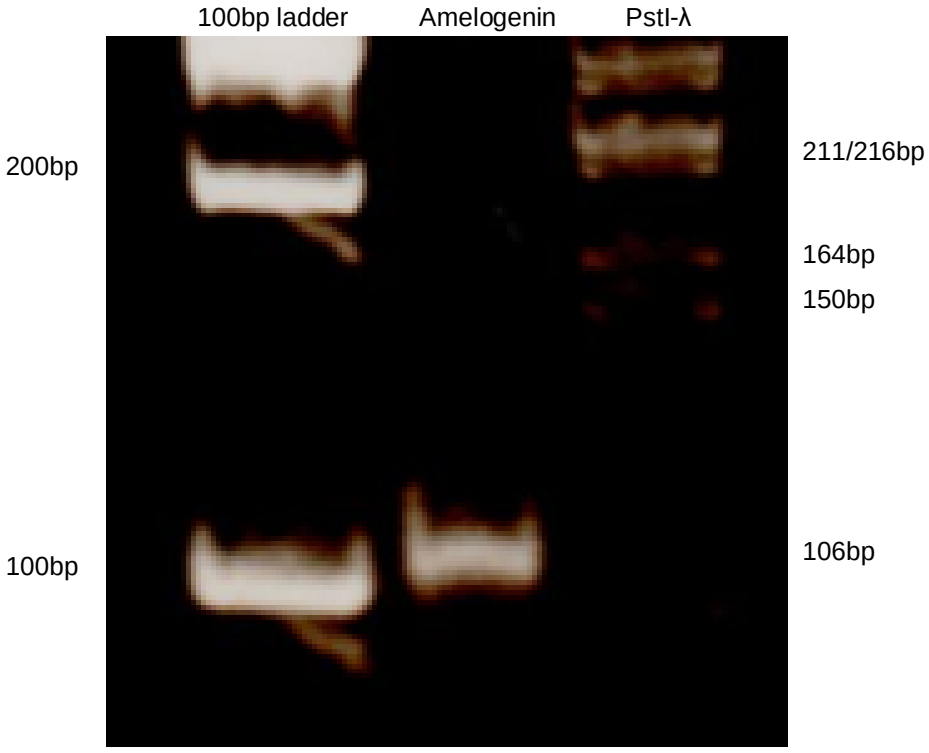

Figure 8: PAGE showing a single band after the amplification for the amelogenin gene, indicating that the specimen was probably female.

2.19: Measurement of DNA concentration using Nanodrop

A Nanodrop (Thermo Scientific) was used whenever the concentration of a DNA product was needed. The Nanodrop is a very quick and efficient way of quantifying DNA when compared to the more traditional method of running the DNA on a gel together with a ladder which has pre-quantified bands. Using the Nanodrop resulted also in higher accuracy. In order to quantify DNA, first a blank reading was taken with the media in which the DNA was eluted. After this three different readings of the same sample were then taken using the Nanodrop and the average of these three readings was taken as the concentration of the DNA.

2.20: Modern male genotyping

The fluorescent dyes used in genotyping are expensive, and thus in order to be able to genotype a large number of STRs in a large number of samples, the technique developed by Schuelke (2000) was used. In this technique instead of the dye being attached to one of the STR-specific primers, the dye was attached to a generic 'tail' primer. When the STR primers were designed, the tail sequence was included at the 5´ end of each of the forward primers. During the PCR, in the first round of cycling, the STR specific primers amplify the target region. After this round the annealing temperature was dropped, which enabled the tail sequence to act as the annealing site, and thus the generic primer with the fluorescent dye attached to it now acts as the amplification primer (Figure 9). This meant that the resulting sequence had the fluorescent dye incorporated in it and could be analysed using the ABI3100 genotyper. Using this technique only four different fluorescent dyes were needed, each with the tail attached, thus reducing the cost of the genotyping considerably.

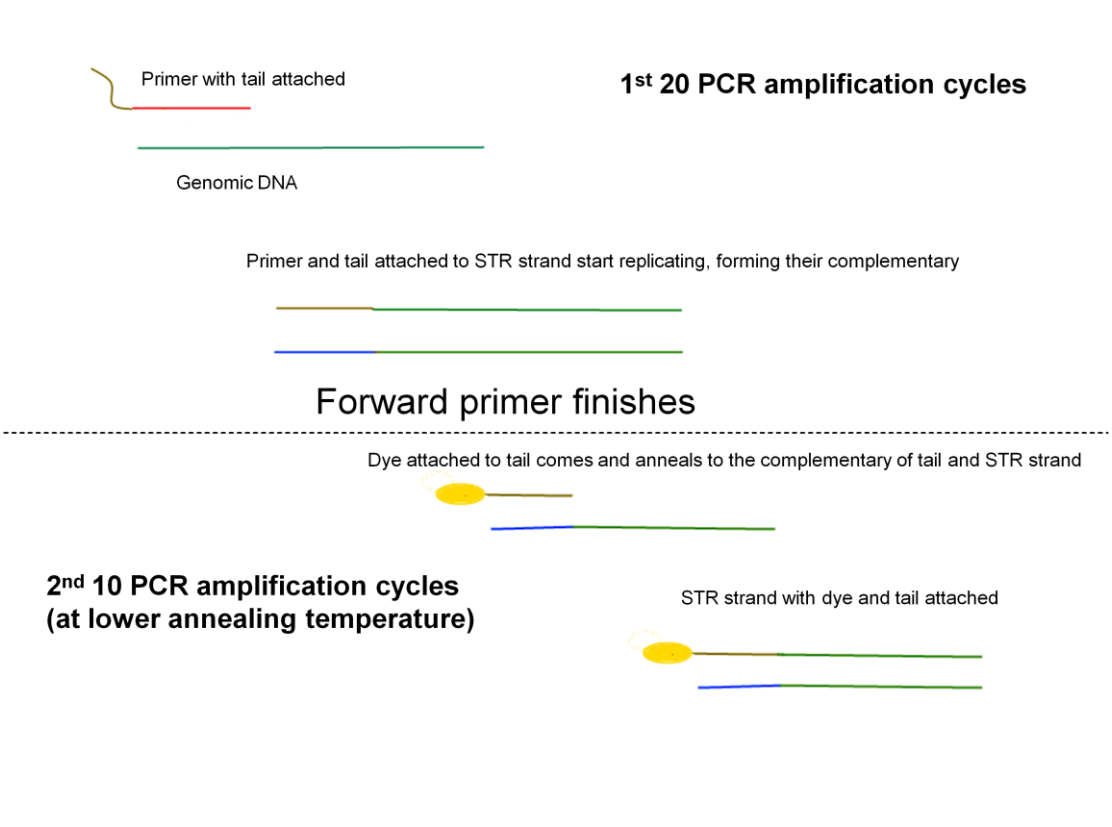

Figure 9: Amplification of an STR by use of a dye attached to a generic forward primer.

In order to screen the results during optimisation the reactions were electrophoresed in a 2% agarose gel and then the PCR was modified until the multiple bands and smearing were reduced or completely eliminated. Smearing normally occurred when the second round of PCR did not 'use up' all of the DNA strands produced during the first cycle, and thus two strands of DNA were present with just the 16bp tail as a difference.

The primers used for the genotyping were taken from various publications (Table 4). Due to the addition of the tail these had to be optimised again as the tail changed the annealing temperatures of the primers (Figure 10). An attempt was also made at multiplexing the reactions but this proved difficult due to the PCR not differentiating between different sequences when attaching the fluorescent dye, since the tail of the sequences was the same one. A higher amount of primer dimer was also observed in the multiplexes due to a higher

amount of unused primers. Some STRs were also amplifying preferentially to others. Thus the attempted multiplexing of these reactions was discarded. A pre-made kit was not used for the genotyping as no commercially available kit incorporated all the STRs that were analysed, another factor of the commercial kits was their prohibitive cost.

Before genotyping, the PCR products from the same sample of certain STRs were mixed together in order to reduce the cost. In the genotyping these were differentiated from each other according to their size and their fluorescent label. Quite a bit of modification was needed ensure that when these STRs were put together their products would not overlap with one another. In order to ensure that the results were reproducible and accurate one of the STRs was genotyped twice independently, using two different fluorescent dyes. STRs of samples which failed to amplify once were repeated and then genotyped with no more than two other STRs. STRs which failed genotyping even in the second round of amplification were discarded.

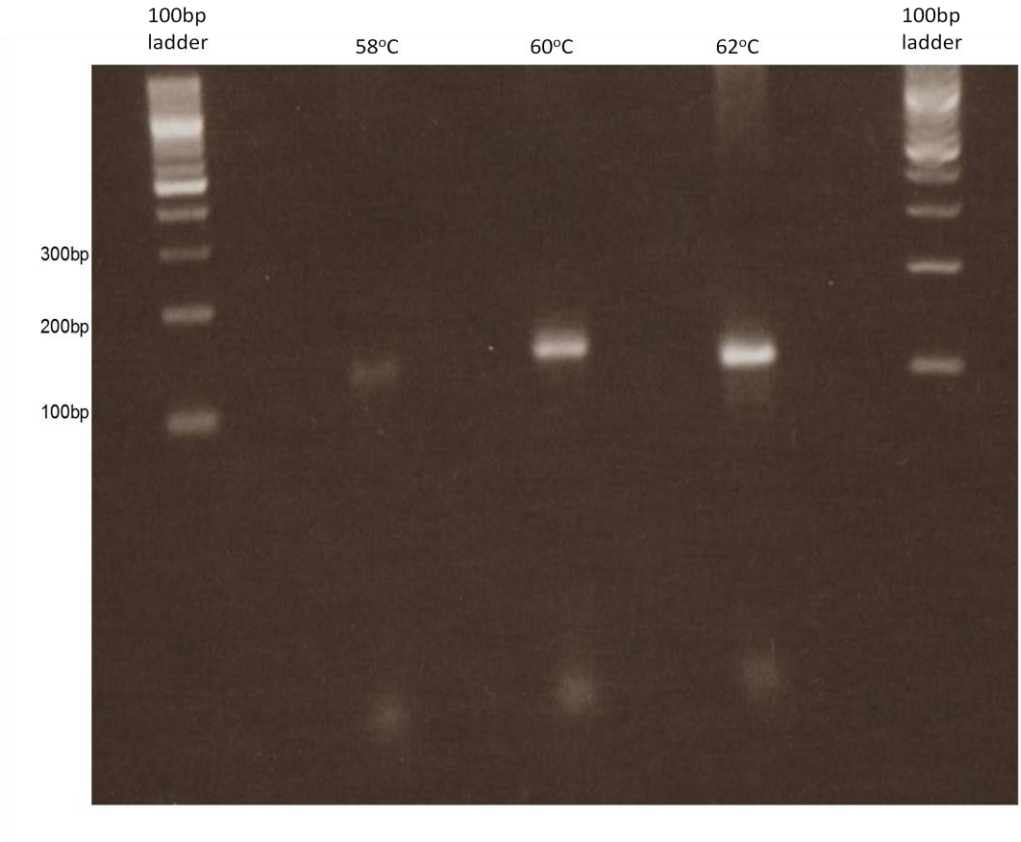

Figure 10: Temperature re-optimisation of an STR after attachment of tail showing that the ideal temperature is 60°C

The genotyping results showed the size of the PCR product together with the tail included. When analysing the data the 16bp of the tail were removed for each individual result, in order to get the true product size and thus allow comparison with the sizes of the products reported after PCR with the original published primers

Name of Primer	Primer Sequence	Reference
DYS456 Forward	5'-GGACCTTGTGATAATGTAAGATAG-3'	Schoske *et al.* (2004)
DYS456 Reverse	5'-GTAGAGGGACAGAACTAATGGAA-3'	
DYS388 Forward	5'-GTGAGTTAGCCGTTTAGCGA-3'	Kayser *et al.* (1997)
DYS388 Reverse	5'-CAGATCGCAACCACTGCG-3'	
DYS460 Forward	5'-CAAATTTGCCAAACTCTTTC-3'	White *et al.* (1999)
DYS460 Reverse	5'-TCTATCCTCTGCCTATCATTTATTA-3'	
DYS390 Forward	5'-TATATTTTACACATTTTTGGGCC-3'	Kayser *et al.* (1997)
DYS390 Reverse	TGACAGTAAAATGAACACATTGC-3'	
DYS391 Forward	5'-CTATTCATTCAATCATACACCCA-3'	Kayser *et al.* (1997)
DYS391 Reverse	5'-GATTCTTTGTGGTGGGTCTG-3'	
DYS393 Forward	5'-GTGGTCTTCTACTTGTGTCAATAC-3'	Kayser *et al.* (1997)
DYS393 Reverse	5'-AACTCAAGTCCAAAAAATGAGG-3'	
Y-GATA-A7.2 Forward	5'-AGGCAGAGGATAGATGATATGGAT-3'	White *et al.* (1999)
Y-GATA-A7.2 Reverse	5'-TTCAGGTAAATCTGTCCAGTAGTGA-3'	
Y-GATA-A4 Forward	5'-TCGACTCGGTACCACTTCCTAGGTTTTC-3'	White *et al.* (1999)
Y-GATA-A4 Reverse	5'-TGGCTTGGAATTCTTTTACCCATCATCT-3'	
DYS447 Forward	5'-GGTCACAGCATGGCTTGGTT-3'	Redd *et al.* (2002)
DYS447 Reverse	5'-GGGCTTGCTTTGCGTTATCT-3'	
DYS385 a/b Forward	5'-AGCATGGGTGACAGAGCTA-3'	Kayser *et al.* (1997)
DYS385 a/b Reverse	5'-TGGGATGCTAGGTAAAGCTG-3'	
DYS392 Forward	5'-AAAAGCCAAGAAGGAAAACAAA-3'	Ruitberg & Butler (2000)
DYS392 Reverse	5'-AAACCTACCAATCCCATTCCTT-3'	
DYS458 Forward	5'-GCAACAGGAATGAAACTCCAAT-3'	Schoske *et al.* (2004)
DYS 458 Reverse	5'-GTTCTGGCATTACAAGCATGAG-3'	

DYS19 Forward	5'-CTACTGAGTTTCTGTTATAGT-3'	Roewer & Epplen (1992)
DYS19 Reverse	5'-ATGGCATGTAGTGAGGACA-3'	
DYS389I/II Forward	5'-CCAACTCTCATCTGTATTATCTAT-3'	Kayser *et al.* (1997)
DYS389I/II Reverse	5'-TCTTATCTCCACCCAGA-3'	
DYS448 Forward	5'-TGTCAAAGAGCTTCAATGGAGA-3'	Redd *et al.* (2002)
DYS 448 Reverse	5'-TCTTCCTTAACGTGAATTTCCTC-3'	
D10S1248 Forward	5'-TTAATGAATTGAACAAATGAGTGAG-3'	Coble & Butler (2005)
D10S1248 Reverse	5'- **G**CAACTCTGGTTGTATTGTCTTCAT-3'	
DYS438 Forward	5'-TGGGGAATAGTTGAACGGTAA-3'	Ruitberg & Butler (2000)
DYS438 Reverse	5'-GGAGGTTGTGGTGAGTCGAG-3'	
DYS437 Forward	5'-GACTATGGGCGTGAGTGCAT-3'	Ayub *et al.* (2000)
DYS437 Reverse	5'-AGACCCTGTCATTCACAGATGA-3'	
DYS439 Forward	5'-TCCTGAATGGTACTTCCTAGGTTT-3'	Ayub *et al.* (2000)
DYS439 Reverse	5'-GCCTGGCTTGGAATTCTTTT-3'	
D18S51 Forward	5'-CAA ACCCGACTACCAGCAAC-3'	Uruquhart *et al.* (1995)
D18S51 Reverse	5'-GAGCCATGTTCATGCCACTG-3'	
Y-GATA-H4 Forward	5'-GAGACCTAAGCAGAGATGTTGGTTTTC-3'	White *et al.* (1999)
Y-GATA-H4 Reverse	5'-CCTCTGATGGTGAAGTAATGGAATTAGA-3'	

Table 4: The original primers which were modified to be used in this study. The sequences given are without the tail attached.

2.21: Data Analysis

The software mtDNA version 3 (http://www.doppiovu3.it/mtdna/) was used as a fast way to change the mitochondrial DNA mutations of samples written as tables in scientific papers into sequences in FASTA format. The parameters used for this program

were to construct the sequence file with the inputted mutations according to the table, from position 16023 to position 16500 of the mitochondrial DNA sequence.

Haplogrep (Kloss-Brandstaetter *et al.* 2010) was the program used for determining haplogroups for Maltese mtDNA sequences. The program uses Phylotree in order to determine the haplogroup of a sequence. The subsequent results where then further checked by the user using Phylotree (Van Oven and Kayser 2009).

Phylotree was used in order to double check the result from Haplogrep, whilst this might seem like circular reasoning since Haplogrep uses the same program to produce its results, Phylotree was a good visual aid to determine the haplogroup of sequences where more than one possibility existed due to a lack of complete HVR2 data .

BioEdit is an alignment editor which was used in order to align different sequences and change sequences into formats desirable for other programs, such as FASTA format and text files which could be edited in Notepad.

GenalEx (Peakall and Smouse 2006) was the program used in order to change the Y-STR alleles into .ych, a format which was accepted by Network for the analysis of the dataset. Pairwise genetic distances were also calculated using GenalEx.

XLSTAT (Addinsoft 2007) was used in order to perform the PCA of the samples when using haplogroup frequencies and when trying to discern patterns in the dataset using individual sequences instead of haplotypes. PCA was conducted in order to view the data in a multi-dimension way.

Arlequin (Excoffier *et al.* 2005) was the program used in order to conduct the Analysis of Molecular Variance (AMOVA) within and amongst populations. This was done in order to see what percentage of variance was there in the populations when compared to the variance amongst populations. The parameters used for this analysis were standard AMOVA computations, locus by locus AMOVA with population specific FST's being computed over 1,000 permutations. Reynold's distance was used in the calculation and the pairwise difference was calculated over 100 permutations at a significance level of 0.05. The results of these analyses were similar to the AMOVA result obtained by using GenAlEx, thus these programs complimented each other in determining that the AMOVA results that were obtained were the correct ones. Haplogroup frequencies were also determined using Arlequin by the 'mere counting' option.

DNAlignment (Fluxus technologies) was the software used in order to change sequences from FASTA format to .rdf files. It also was used as an alternative, faster sequence aligner to BioEdit in large datasets although it lacks some of the functionality of BioEdit. The need of the file format change was due to Network (Fluxus technologies) not accepting FASTA files, and thus another program had to be used in order to change the sequences to a format acceptable to Network.

Network was used in order to construct reduced median networks for the mtDNA data and median joining networks for the Y-STR data. Reduced median networks were used to reconstruct phylogenies which show relationships between different samples, whilst at the same time resolving likely parallel mutational events and producing reticulations in order to show where conflict still exists (Richards *et al.* 1996). Reduced median

networks are a good way to infer homoplasy. The parameters used for the mtDNA data analysis were:

The weighting for positions 16129, 16189, 16311, 16362 was 2 whilst the weighting for 16093, 16172, 16209, 16223, 16278, 16293, 16294 and 16325 was 5. All other positions were weighted at 10. This was done in order to make up for the fast mutation rates some positions have when compared to others which created reticulations in some networks. This method was used by Richards *et al.* (1998) and reduces the reticulation problem. In the current study the networks for haplogroups H, T, R, U, L, M and K were weighted due to the presence of a large number of reticulations, whilst the networks for J, W and HV were not weighted as this problem did not occur in them.

The Lebanese mtDNA dataset (Haber *et al.* 2012) contains only the controls of the study and not the full dataset including the patients with coronary artery disease. This was done in order to make sure that a random dataset was used for the population without any bias towards specific haplogroups since the study showed that some haplogroups showed a significant relationship with coronary heart disease.

For the 8 Y-STR data one of the shortest spanning trees was shown as the dataset was too big to accurately interpret otherwise. Another problem with the dataset when using 8 Y-STRs in seven populations was that the number of nodes formed was greater than the limit of the program, thus in order to obtain results, the Frequency >1 criterion was used. Thus, all the visible combinations in the Network were found at least twice in the dataset, except for the Maltese samples, where all the samples are shown in the dataset and not just the combinations which two or more samples had.

The same problem occurred with the 5 Y-STR analysis, where due to the larger number of samples, the analysis had to be divided between the Western, and the Eastern Mediterranean, with the Maltese samples being compared to both datasets. All samples except for the Maltese ones were also analysed using the Frequency>1 criterion.

The program was also used in an attempt to date clusters, but unfortunately this attempt was not very successful due to the nature of the data being analysed which meant that events that happened in history were genetically indistinguishable from those happening in Maltese prehistory. The parameters used for this analysis were 1 substitution every 16677 years between positions 16051 and 16400 excluding positions 16182C, 16183C and 16194C as listed in the paper by Soares *et al.* (2009).

DnaSP (Librado and Rozas 2009) was used in the study in order to estimate nucleotide diversity (π), using the Jukes and Cantor model. The program was also very useful in converting the mtDNA modern population datasets into Arlequin format.

Chapter 3: Ancient DNA analysis

3.1: Objectives of the aDNA study and the problems encountered during its excecution

Ancient DNA analysis was conducted in order to see whether kinship relationships could be determined between samples from the same site, and also in order to identify any relationship different sites might have in the Maltese archaeological record. The study was also conducted in order to provide information about aDNA survival in the Maltese archaeological record and to try and set a limit after which aDNA analysis in the Maltese islands are might not be deemed to be feasable currently using the methods in this study. A further aim of the study was to compare the aDNA sequences with modern DNA sequences in order to determine if any of the polymorphisms found in the archaeological samples can still be found in the modern Maltese population or other populations from around the Mediterranean which might have had a genetic impact on Malta during the particular time of the samples.

A problem which was encountered in the aDNA study was that whilst aDNA sequences were amplified, their replication success rate left much to be desired. Unfortunately no sample yielded the whole of the HVR1 and 2 sequences. Another factor which hindered the study was contamination in the samples. Whilst some samples yielded known contaminants and thus were removed, others yielded two different kinds of polymorphisms in separate clones, none of which could be attributed to the researchers or archaeologists who came in contact with the sample. Of hindrance to the study was also the high sensitivity of GelRed,

which would show up aDNA bands which were at such a low concentration that they could not be succesfully cloned.

3.2: Ancient DNA amplification and interpretation

Ancient DNA analysis was conducted as described in Chapter 2. The results are summarised in Table 5. The sequences obtained from the study can be found in Appendix 4. The sequences which were in agreement with the CRS are included even though some of the archaelogists who excavated the site had this sequence, the reason being that haplogroup H is the most common haplogroup in Europe and thus it is not easy to distinguish whether the DNA is ancient or modern when the CRS is obtained.

Haplogroup identification was done by taking into account the haplogroups which contained the amplified polymorphisms but excluding the ones which had other polymorphisms listed in the amplified region of the aDNA but which were not visible in the sample. This process was done using the program Haplogrep.

Sample	AA	DSO	NC	1st EX	2nd EX	Region amplified	Polymorphisms in amplification	Probable Haplogroups	Modern contamination?
BR1	10	3	12	Yes		16209-16340	16234T-16247G	U8b/U9b/U2c	No
						16209-16340	16264T-16291T	H2a5a	No
					No	16147-16294	16189C-16223T	X, M, U4	No
				Yes	No	Amelogenin	Probable Female		No
BR2	8	0		No	No	N/A			
BR3	8	0		No	No	N/A			
BR4	8	2	15	Yes		16209-16340	16264T-16278T	H, R, J, U	No
					Yes		CRS	H	Contamination
BR5	8	0		No	No	N/A			
BR6	8	0		No	No	N/A			
BRF2S1	8	0		No	No	N/A			
BRF2S2	8	0		No	No	N/A			
BRF2S3	8	0		No	No	N/A			
BRF2S4	8	1	4	Yes		16147-16294	16183C-16189C-16241T-16270T	U5b1	No
					No	N/A			
TG26	8	3	12	Yes		16147-16294	CRS	H	Contamination
					Yes		CRS	H	Contamination
				Yes		16209-16340	16261T	H	No
					No				
TG32	8	2	27	Yes		16209-16340	16256T-16301T-16311C	R0b, H1, J1c2a, U2e1, H2	No
					Yes		16256T-16301T-16311C	R0b, H1, J1c2a, U2e1, H2	No
TG33	11	4	36	Yes		16209-16340	CRS	H	Contamination
					Yes		16282A	H1as1a	No
				Yes		16147-16294	16247G	H	No
					Yes		CRS	H	Contamination
TG36	8			No					
					No				
TG115A	8	2	7	Yes		16209-16340	16294T	T, H42	Contamination
							16294T	T, H42	Contamination
					Yes		CRS	H	Contamination
TG115B	8	4	16	Yes		16146-16294	16172C	U6a	No
					No	N/A			
				Yes		16209-16340	16278T	U6a	No
					Yes	N/A	16278T	U6a	No
							16278T-16286T-16295T	U6a	No
				Yes		16270-16420	16362C-16399G	U6a	No
					No	N/A			
TG115C	8	2	13	Yes		16209-16340	16234T	H1	No
							16256T-16301T-16311C	R0b, H1, J1c2a, U2e1, H2	No
							16234T	H1	No
					Yes	16209-16340	16256T-16301T-16311C	R0b, H1, J1c2a, U2e1, H2	No
K18	8	1	1	Yes		16-132	73G	H1a, R, JT, J, T, F, B, U, K, X, L, M	Unknown
					No	N/A			
K29	8	1	4	Yes		16147-16294	16187T-16188T-16190T-16242T-16245T	L0a2, H, D4c	Unknown
					No	N/A			
K32	8	0							

Table 5: List of Ancient DNA samples and the regions amplified. AA = Amplification attempts, DSO = Different sequences obtained, NC = number of colonies, 1st EX = first extraction, 2nd EX = second extraction from separate sample. The alternating highlighted parts indicate a different inhumation.

The only nuclear DNA which was successfully amplified was from the amelogenin gene. This was successfully amplified in only one sample, BR1, and the resulting gel showed that the sample was probably female. Unfortunately this result could not be replicated. All other autosomal and Y chomosome amplifications failed. The probable reason for this is the higher amount of mtDNA that is present in a cell when compared to nuclear DNA, thus rendering the amplification of mtDNA more probable.

The study was a success since although not all the samples which were tested yielded aDNA, the amplification of ancient DNA from some samples shows that it is able to survive in the Maltese archaeological record, this success is counteracted by the amount of samples which yielded mixed sequencies, thus indicating that a large amount of samples have suffered from contamination.

As can be seen from Table 5 contamination was a big issue in the aDNA analysis of samples from Malta. Thus from 11 samples that yielded sequences from the analysis only three sequences (TG115B, TG32, BrF2S4) yielded sequences which were not mixed and did not have a possible source of contamination amongst the people who came in contact with the sample. Samples Tg115C and BR1 yielded mixed sequences but without any of the haplotypes being found in the possible list of contaminants. On the other hand samples TG115C, TG33, TG26, and BR4 yielded mixed sequences where at least one of the sequences could be traced to a modern contaminant. The Kercem samples due to their low clone yield and their lack of replication cannot be classified as beeing real aDNA or contaminants.

3.3: Sequence analysis of the Tal-Gardina samples

TG115B is one of the samples which most strongly suggests that aDNA survives in the Maltese archaeological record. The sample can be defined as probably haplogroup H with the mutation motif not found in the Modern Maltese population. TG115C has mixed sequences which are not found in the possible lists of contaminants which came in contact with the samples. This might indicate that the contamination is ancient, although the possibility remains that the contamination is modern from a person which has not been tested in this study, even though the maximum effort was made to trace all persons who may have come in contact with the inhumations and test them.

TG115A yielded two sequences, one of which was the CRS and the other one displayed mutation 16294 which can be part of haplogroup T. Unfortunately the researcher who conducted the analysis has also got the same SNP and thus it can be concluded that this sequence might be contamination from the researcher doing the study. The nature of aDNA analysis means that even if this sequence is truly aDNA the possibility of contamination cannot be excluded unless another researcher with a different haplogroup does the analysis. On the other hand this does not mean that the CRS haplotype is the true ancient sequence since this sequence is very common in modern day Europeans and it could be the result of contamination of the sample.

TG32 is another sample which strongly suggests that aDNA has survived in the Maltese archaeological record, as both the original extraction and all subsequent ones yield the same polymorphisms. Unfortunately the haplogroup of the sample could not be assigned as a variety of haplogroups have at least one the SNP found in this sample included in their

mutation motifs. The haplotype of this sample is not found in the modern Maltese population.

TG33 is a problematic sample as all the sequences which were amplified from this sample can be part of haplogroup H, but the difference between ancient DNA, modern contamination or degradation in the sample cannot be clearely defined although the presence of CRS in all the successful amplifications indicated that this is most probably the aDNA sequence, unless a modern contaminant managed to contaminate different parts of the same inhumation.

TG26 yielded two sequences, one was CRS which is found both in the modern population and in the list of possible contaminants. The other sequence showing a SNP at position 16261 is found in the modern Maltese population but not in the list of the possible contaminants.

3.4: Sequence analysis of the Tal-Barrani samples

The Tal-Barrani samples had a lower replication success rate than the Tal-Gardina site, even though they come roughly from the same period. The probable reason for this is that the samples were excavated seventeen years before the DNA analysis, compared to two years for the Tal-Gardina site. The accelerated rate of degradation and the lower rate of replication success after samples have been excavated has been described by Burger *et al.* (1999) and this study supports his study in this respect.

BR1 yielded multiple sequences which are not found in the list of contaminants and thus the true aDNA haplotype of the sample could not be determined. BR4 yielded two sequences, both CRS and a haplotype showing a SNP at position 16278 and another in some samples at position 16264 was found in the first extraction, whilst in the second extraction only the sequence with the SNP at positions 16278 and 16264 was found, thus indicating that this might be the true aDNA sequence.

BRF2S4 yielded a single haplotype which indicates that it forms part of haplogroup U5. This haplogroup indicates that this sample was European since that haplogroup is thought to have originated from Europe. Unfortunately the result could not be replicated. Of interest is that this haplogroup is not found in the modern Maltese population.

The result suggests that there was no difference between the haplogroup composition of the rural and the urban parts of Malta during the Roman period, with both having both European and African haplogroups. This also reflects the empire at that period of time, where Roman citizenship was granted not just to the inhabitants of Rome and the Lazio region, as in the time of the republic, but to citizens and cities from all around the empire. The site also produced the only aDNA sex identification of the study. BR1 was identified as being probably female. The result could not be replicated.

Although the results are limited in number, the samples taken from the two features of Tal-Barrani do not seem to show any difference in aDNA preservation, even though the samples from feature two were thought to have been buried at an earlier date than those of feature one. This is a further indication that the burial conditions are more relevant to aDNA survival than the age of the burials. Unfortunately the large extent of contamination found in

the samples indicate that better safeguards are needed in Maltese archaeology whilst excavating sites if aDNA studies are to be successful.

3.5: Sequence analysis and conclusions derived from the Kercem samples

The Kercem samples were the oldest samples in the Maltese archaeological record to yield DNA results, although doubt remains about whether the results are modern or ancient. Only two out of the three Kercem inhumations yielded any results. The two results came from the bone samples taken from the inhumations and not from the teeth samples.

Sample K29 yielded polymorphisms which could not be attributed to any haplogroup. This might mean either that more of the HVR1 region would need to be amplified in order to determine the haplogroup, or that the haplogroup the sample belongs to has died out in history. The polymorphisms in this sample are not present in the modern Maltese population. This might be an indication that the people living on the Maltese islands at that time were replaced in prehistory. Unfortunately, since the result was not succesfully replicated, no definite conclusions can be made, although the fact that the polymorphisms are not found in the modern Maltese indicates that the amplified DNA is ancient. A problem with this sample is that some of the SNPs found in it look like artifacts which puts any claims made here about the sample in doubt.

The only other Kercem sample that yielded results was K18 but unfortunately this result was from a region from HVR2 and thus not as informative. The only conclusion that can be made from this result is that the sample did not have an H haplogroup due to the polymorphism at position 73. The result could not be replicated, probably due to the age of

the samples, as the Kercem site was the oldest Maltese site in which aDNA analysis was conducted. Due to the presence of this polymorphism in the modern Maltese population the argument for the amplified DNA being ancient cannot be put forward as with K29.

3.6: Comparison of aDNA samples with modern Western Mediterranean populations

A median joining network was constructed comparing the aDNA sequences that were obtained from positions 16209 to 16293 of HVR1 with the equivalent sequences of the other populations from the Mediterranean region (Figure 11).

The results show that most of the ancient Maltese sequences form a cluster with modern Maltese samples or are one mutation away. The exceptions are; two time were they are one mutation away from a cluster containing various regions from the Mediterranean, the Kercem sample which is one mutation away from a Greek sample and a Tal-Barrani sample which is one mutation away from a cluster containing a Tal-Gardina sample and a North Italian sample.

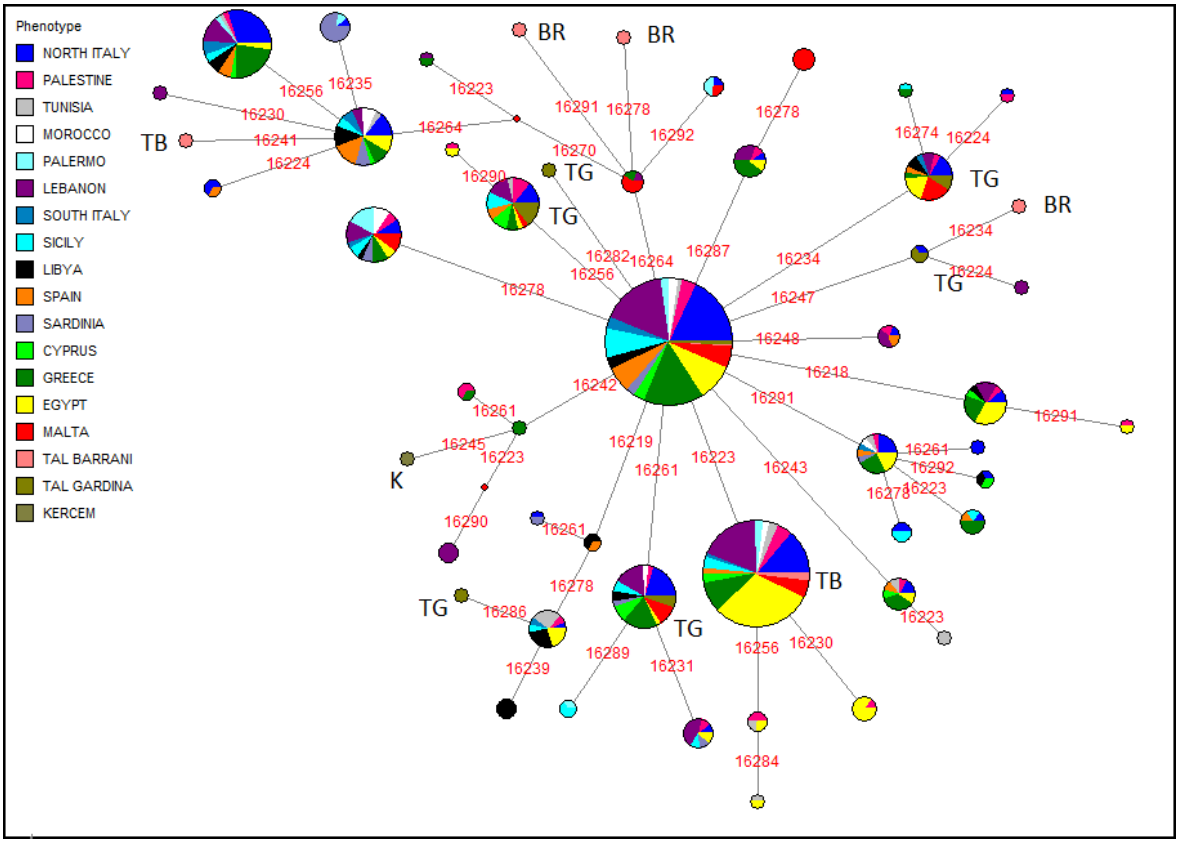

Figure 11: A reduced median joining network tree constructed from an 84bp region of HVR1 from modern and ancient samples. Tal-Gardina = TG, Tal-Barrani = BR, Kercem = K.

The clustering of most aDNA samples with clusters containing modern Maltese samples might be an indication that some of the haplotypes present in late Roman Malta are possibly still present in the Maltese population today, or that the people who repopulated Malta were of a similar genetic stock as the people who lived there during the late Roman period. Of interest is that a Tal-Gardina sample clusters exclusively with a North Italian sample, and this yields another sequence one mutation away composed of a Tal-Barrani sample.

Caution must be taken when making conclusions from this type of analysis since the network is only being used to try and visualise patterns in the sequences as clusters hail from multiple haplogroups which might be represented in more than one cluster due to the short lenght of the region under analysis. Another word of caution is that only a small segment of the HVR1 has been analysed since no aDNA sample yielded the whole HVR1 region, any

patterns seen here might not necessarily reflect reality. The warning is especially useful when haplogroups U5 and U6 are kept in mind, where although they are seen in the aDNA sequences, they are not found in the modern population.

3.7: Conclusions from the aDNA study

The aDNA study has shown for the first time that aDNA survives in the Maltese archaeological record, although the quality of the DNA and the amount of contamination in this study means that further studies are necessary to ascertain this. The Maltese from the Roman period and their probable haplogroups show signs of coming both from the European continent and from the African continent, which is reflected by the geographic reality of Malta being in the middle of 'Mare Nostrum'.

Contamination was also a problem in the analysis. Whilst established precautions were taken in the extraction and amplification of the DNA, contaminant sequences were obtained when some samples gave mixtures of sequences which could not be traced back to the excavators or the researchers working on the samples. Another issue was haplogroup H, since this is the most common haplogroup found in Europe, some of the archaeologists working on the samples had this haplogroup, and thus the aDNA amplifications which showed this haplogroup, even though they may have been truly ancient, could not be proven to be so.

Chapter 4: MtDNA of the Maltese population.

4.1: MtDNA diversity of the modern Maltese population

MtDNA analysis was conducted in order to determine the Maltese matrilineal heritage and whether this varied from the paternal heritage in terms of admixture with other populations from the region. This was done by amplifying HVR1 and 2 of samples of the Maltese population from primers which were designed and optimised for this study. Of a total of 144 samples, 133 yielded both HVR1 and position 73 of HVR2 after two amplification attempts on each. A table was then constructed showing all the different haplotypes present in the Maltese sample (Appendix 5). The results of the Maltese samples were then compared to other studies conducted on populations from around the Mediterranean region, in order to see which populations are the most similar to the Maltese one matrilinealy.

4.2: Identity and frequency of haplogroups present in the Maltese population

Most of the haplogroups in the Maltese population were typical European haplogroup lineages, but haplogroups associated with Africa were also present in the sample, the most prominent of these being haplogroup L2 which is found in 11% of the modern Maltese population. Haplotypes which are found at a low frequency in Europe were also found at a low frequency in Malta, amongst these was haplogroup M. The samples listed as haplogroup JT could not be identified as either of J or T even when a reduced median Network with all Maltese haplotypes was drawn (Figure 12) in order to search for the sequences' neighbours, as suggested by Bandelt (*et al.* 2001) and also after the haplotypes were looked for on

EMPOP (Parson and Dür 2007) in an attempt to find similar sequences. The sequences were finally determined to be R0a.

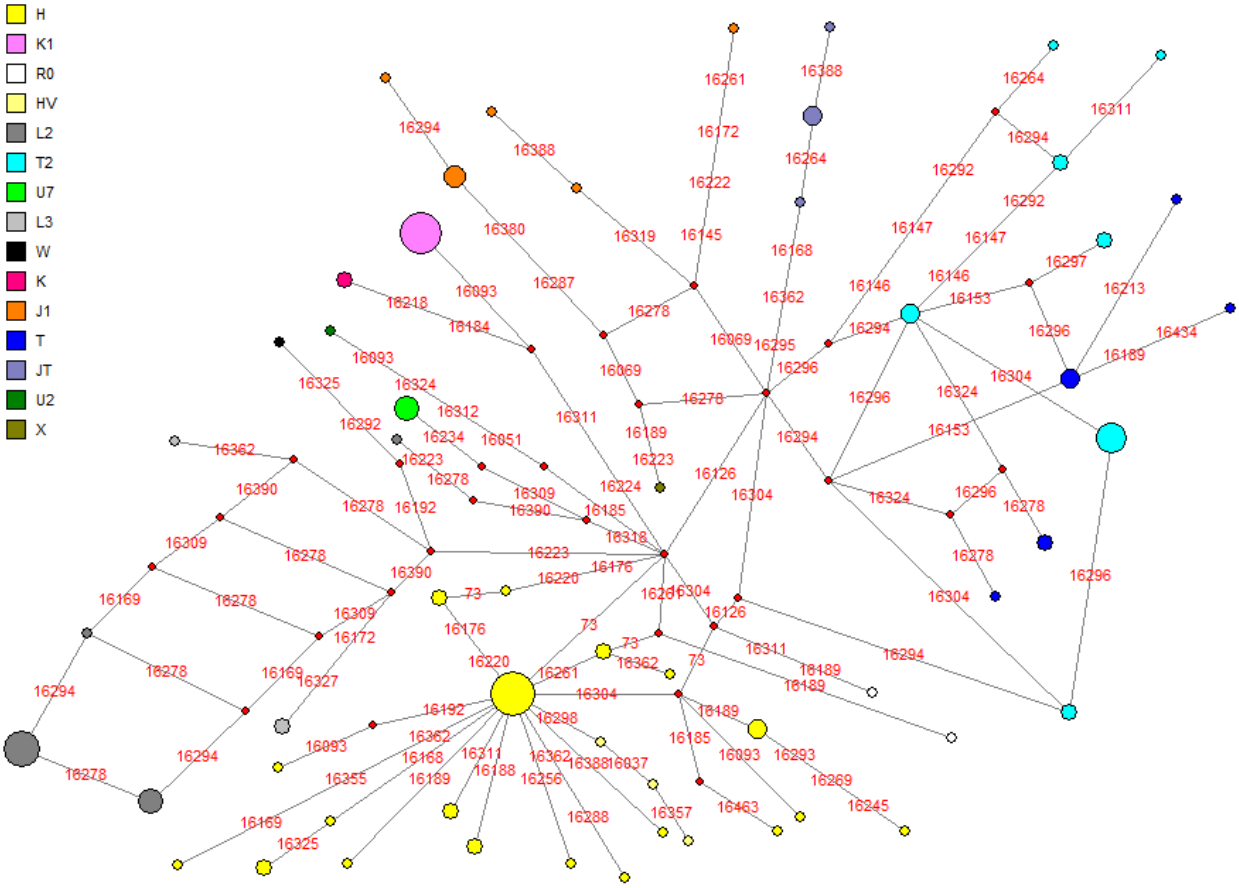

Figure 12: Reduced Median Network of Maltese samples and the respective Haplogroups assigned to them after comparison to Haplogrep and Phylotree.

4.3: General relationships between the Maltese and other populations

In order to see the nucleotide diversity in populations, π was calculated using the Jukes and Cantor method, which assumes that all changes are equally likely using the software DNAsp. The results showed that Malta was in the middle of the populations with regards to nucleotide diversity, where the Maltese samples were not as diverse as the Egyptian and Tunisian ones, but they were not as highly conserved as the Spanish and Palermo samples.

The results for the Palermo population could also be affected by its small population sample size.

Population	π	Standard Deviation	Reference	Size
Spain	0.00993	0.00071	Prieto *et al.* (2011)	116
Palermo	0.01078	0.0011	Vona *et al.* (2001)	47
North Italy	0.01161	0.00038	Turchi *et al.* (2008)	395
Sardinia	0.01169	0.00094	Fraumene *et al.* (2003)	63
Sicily	0.01283	0.00065	Ottoni *et al.* (2009)	154
Greece	0.01341	0.00049	Irwin *et al.* (2008)	319
Lebanon	0.01374	0.00044	Haber *et al.* (2012)	363
Malta	0.01436	0.00065	This study	132
Cyprus	0.01462	0.00107	Irwin *et al.* (2008)	91
Palestine	0.01517	0.00076	Richards *et al.* (2000)	115
Morocco	0.01542	0.00145	Turchi *et al.* (2009)	56
South Italy	0.01623	0.00081	Ottoni *et al.* (2009)	187
Libya	0.01657	0.00062	Fadhlaoui-Zid *et al.* (2011)	263
Egypt	0.01738	0.00065	Saunier *et al.* (2009)	273
Tunisia	0.02020	0.00131	Turchi *et al.* (2009)	64

Table 6: Nucleotide diversity of different populations from around the Mediterranean.

4.4: The relationship of different populations from around the Mediterranean as determined by principal component analysis (PCA) and AMOVA

4.4.1: AMOVA and PCA of populations from around the Mediterranean region

When an analysis of molecular variance (AMOVA) of populations from around the Mediterranean was performed (Figure 13) using Arlequin (Excoffier *et al.* 2005), the results showed that the variation amongst populations was 98%, compared to the 2% of variation between different populations.

```
--------------------------
AMOVA design and results :
--------------------------

           Weir, B.S. and Cockerham, C.C. 1984.
           Excoffier, L., Smouse, P., and Quattro, J. 1992.
           Weir, B. S., 1996.
```

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	14	163.440	0.05996 Va	2.19
Within populations	2326	6243.313	2.68414 Vb	97.81
Total	2340	6406.752	2.74411	

Fixation Index FST : 0.0218

Figure 13: mtDNA HVR1 AMOVA result of populations from around the Mediterranean region.

In order to make sure that this did not mean that the observed variation was random, a frequency distribution of random variance vs observed variance was calculated using GenAlEx (Figure 14), this analysis showed that the variance is not random as the pbserved variance is significantly different to the variance range which had been generated randomly.

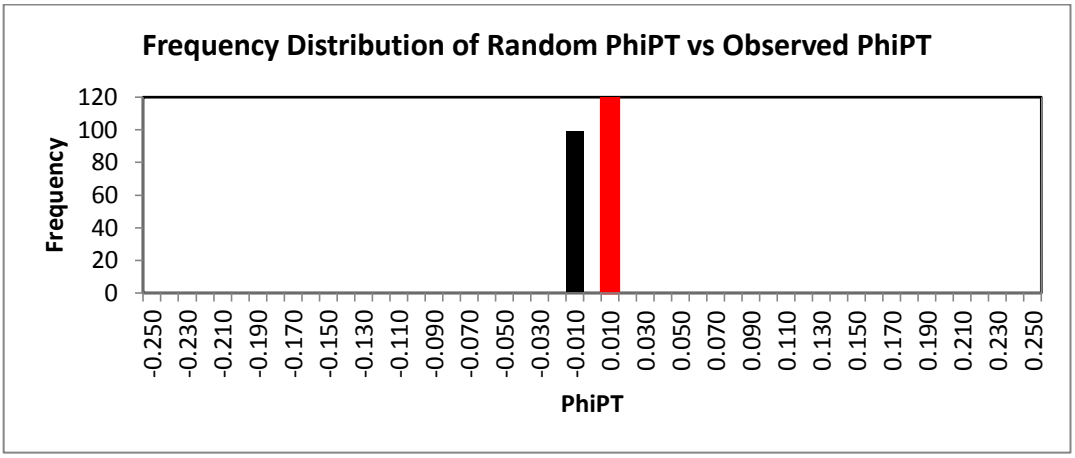

Figure 14: Difference between the random variance and the observed variance over 999 permutations.

Principal component analysis was conducting using major haplogroup frequencies of fifteen populations from around the Mediterranean (Appendix 6). The result from PC 1 vs PC 2 which has the highest variance showed the Maltese population coming closest to the South Italian, Palestinian, Sicilian and the samples from Palermo (Figure 15). The PCA also shows a geographic divide that places the populations coming from the North of the Mediterranean to the left, and the populations coming from the South of the Mediterranean to the South. The exception to this is the Palestinian population which clusters with the Sicilian samples.

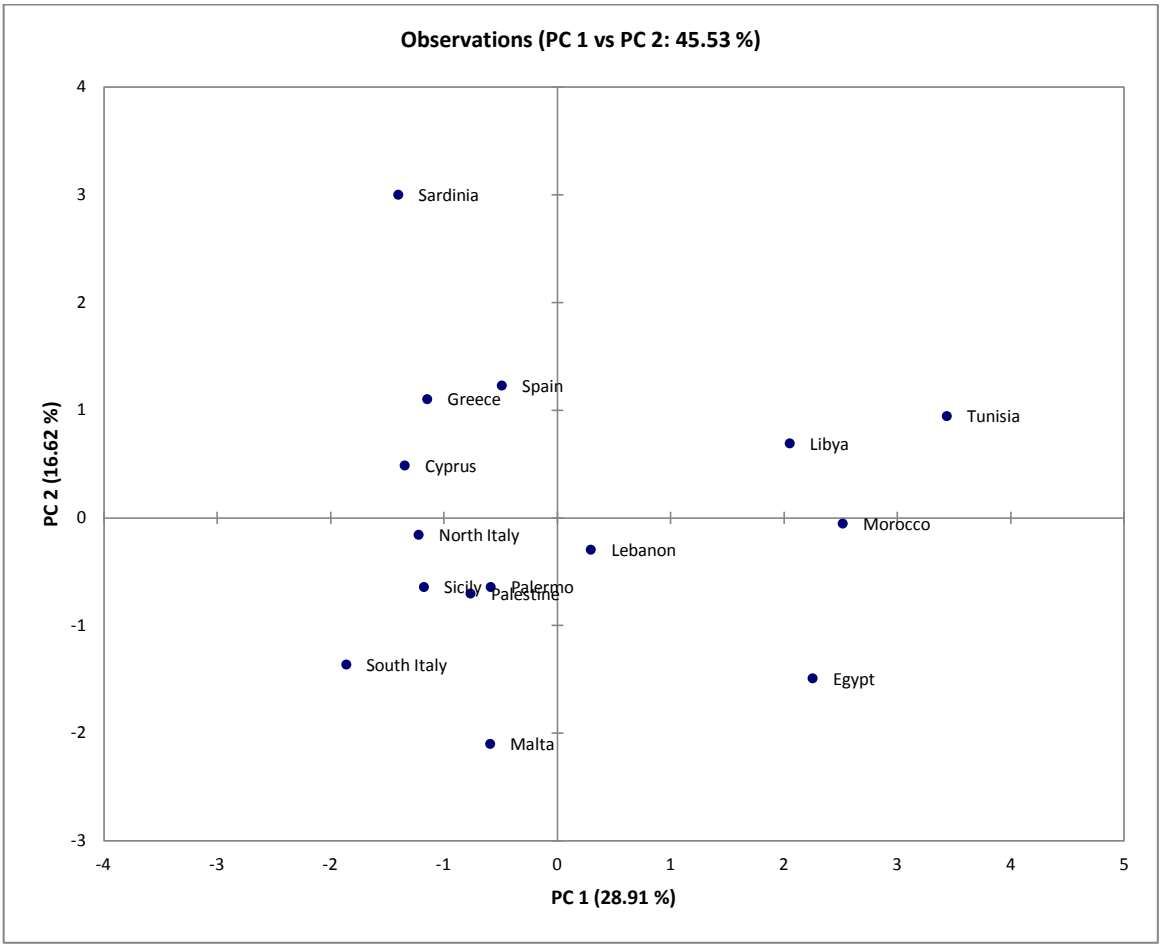

Figure 15: PC 1 vs PC 2 of 15 populations from around the Mediterranean.

The analysis of PC 1 vs PC 3 (Figure 16) is as important as the previous one since it covers 43.64% of the variability. In this analysis Malta places close to the Sicilian, North Italian and Lebanese populations with the South Italian, Cypriot and Spanish populations following. The geographic divide is also visible in this analysis with the Northern samples clustering to the left and the Southern Mediterranean samples clustering to the right. The Palestinian population is still the only population which breaks this pattern when it positions itself closest to the Greek and Sardinian samples further to the left than the Maltese and Spanish samples.

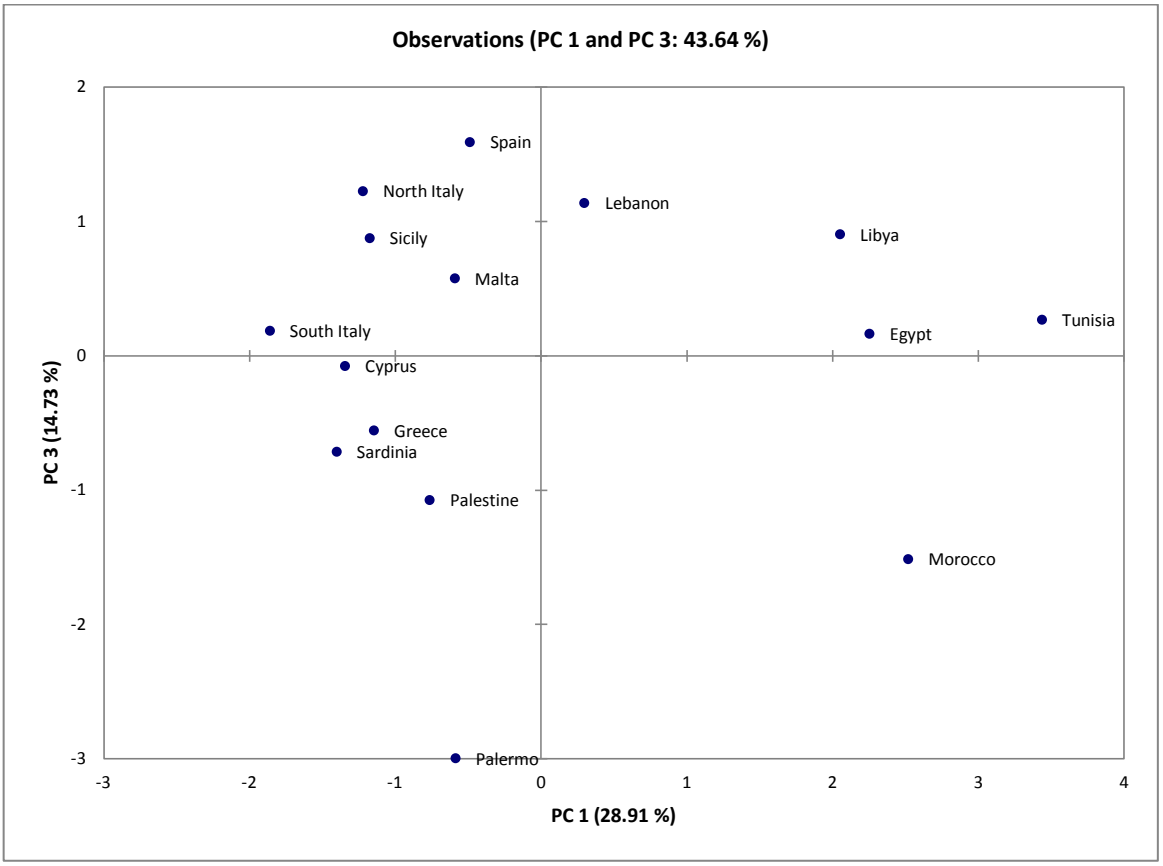

Figure 16: PC 1 vs PC 3.

In PC 2 vs PC 3 (Figure 17) as variation decreases the geographical pattern is not present anymore and the samples seem to cluster randomly with no geographical affiliation. Thus Lebanon has got the North of Italy as its closest neighbour, and Egypt clusters with South Italy. Malta in this analysis has got Egypt, South Italy and Sicily as its closest neighbours.

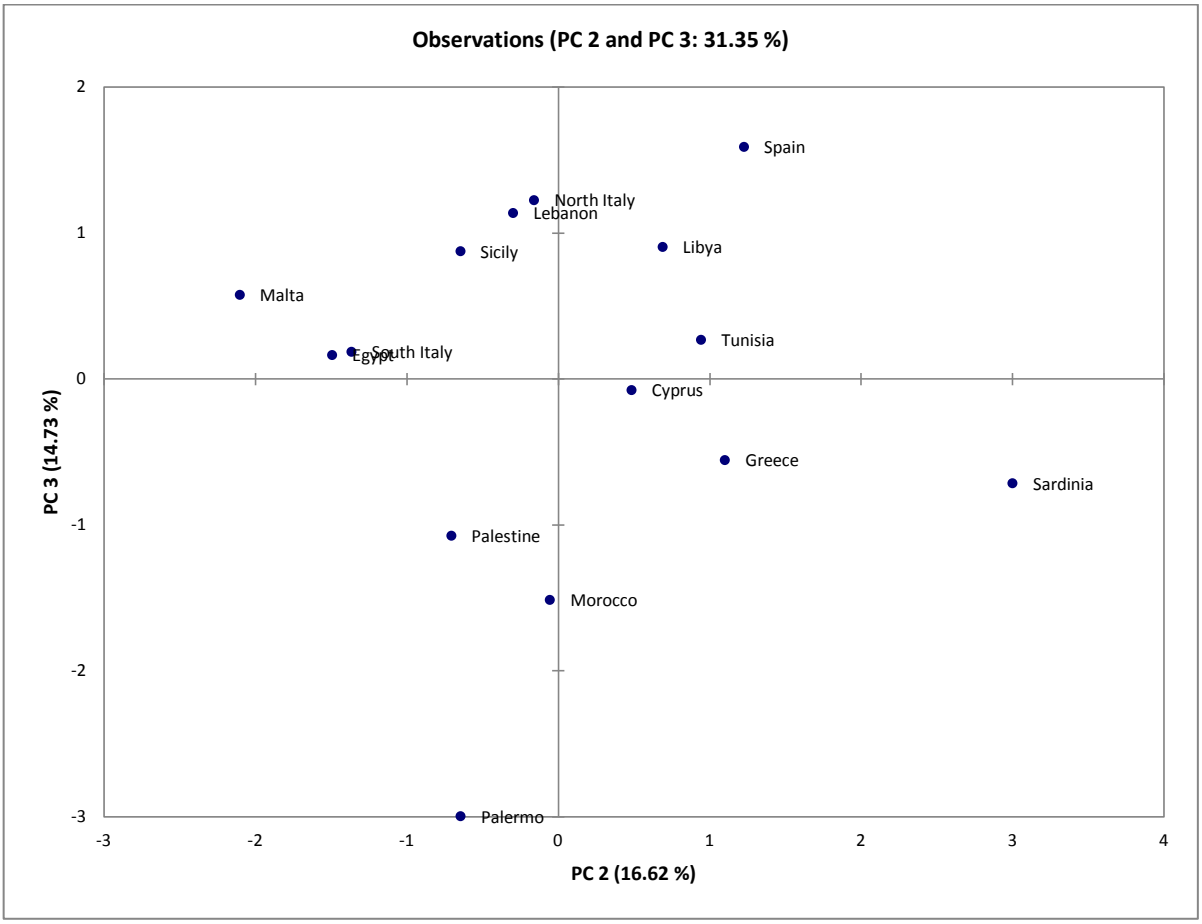

Figure 17: PC 2 vs PC 3 showing a breakdown in geographical affiliation with Malta being closest to Egypt and South Italy.

In all the PC plots the Maltese population always had the Sicilian or the South Italian population as one of its neighbours, even if they were not always the closest. The other populations such as the Egyptian and the Palestinian populations, although they are closest to the Maltese population in some of the plots do not maintain their close proximity through all the plots as the Sicilian and South Italian samples do.

4.4.2: PCA from information index analysis of individual nucleotides

An attempt was made to analyse the sequences not in terms of haplogroups, but in terms of individual nucleotides. In order to do this a PCA was conducted using the diversity information index of each nucleotide in respect to the population. The diversity information index was calculated using GenAlEx.

The variation obtained in the PCA using this method was not as high as desired, this was probably due to the large number of nucleotides which were analysed and the similarity of the samples being analysed since they all come from the same species.

In the first PCA which reaches 36.66% variability (Figure 18) a large number of samples cluster near each other. A closer look at these populations (Figure 19) reveals that the Maltese population is closest to the samples from Palermo, Libya, Palestine, and South Italy, with Sardinia and Sicily following close after.

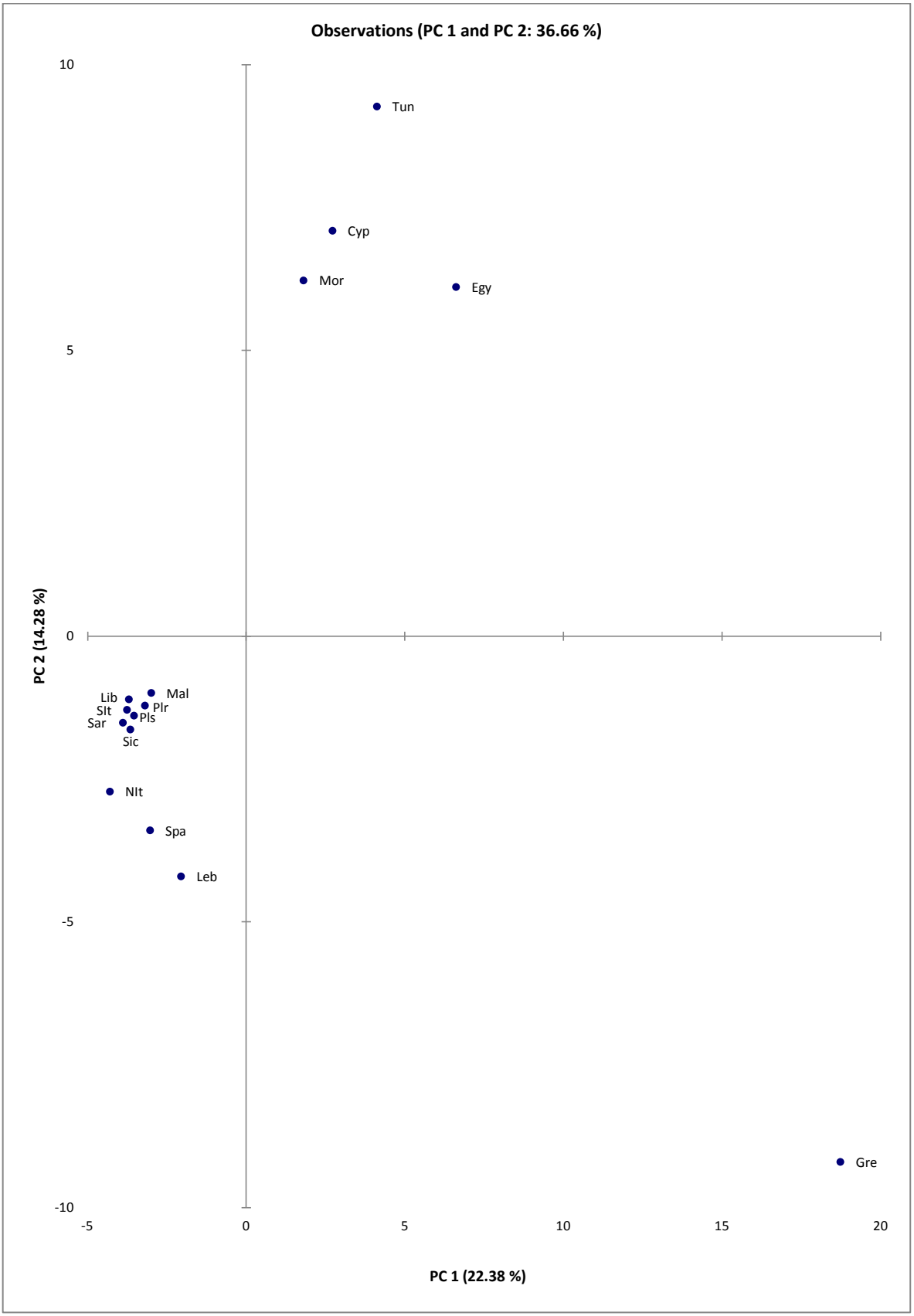

Figure 18: PC 1 vs PC 2 of individual polymorphism analysis. Tun: Tunisia, Cyp: Cyprus, Mor: Morocco, Egy: Egypt, Mal: Malta, Plr: Palermo, Lib: Libya, SIt: South Italy, Pls: Palestine, Sar: Sardinia, Sic: Sicily, NIt: North Italy, Spa: Spain, Leb: Lebanon, Gre: Greece.

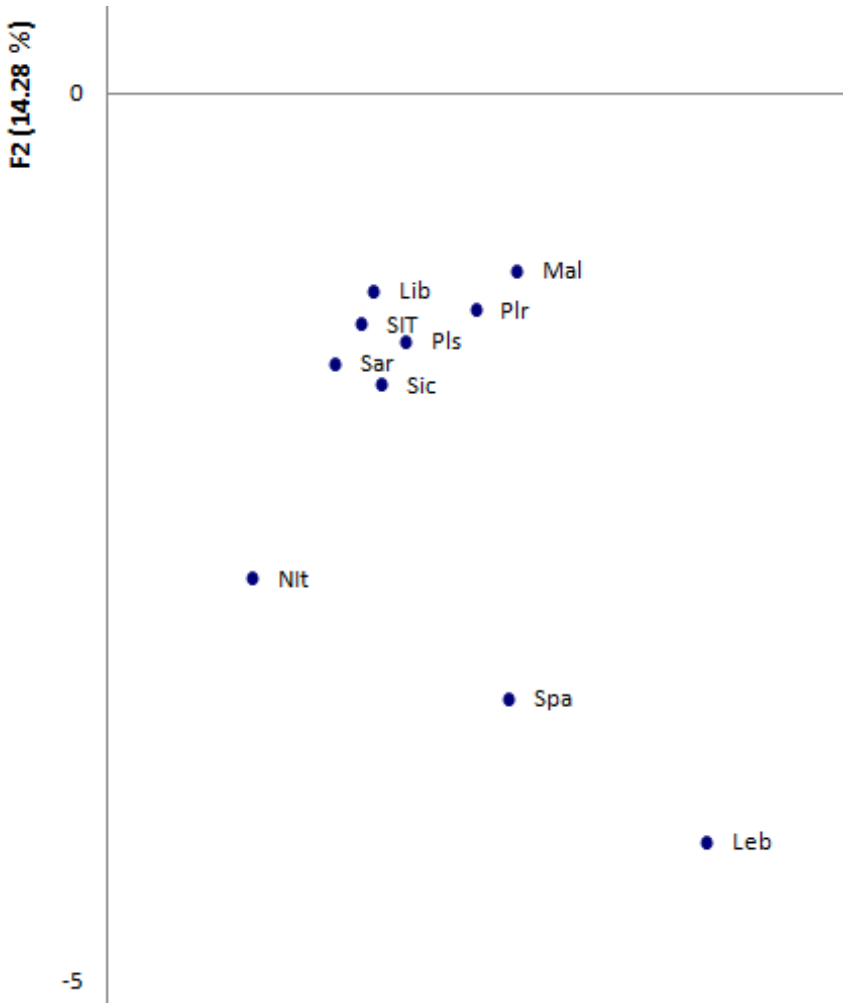

Figure 19: Inset of PC 1 vs PC 2.

PC 1 vs PC 3 were then analysed (Figure 20) and the PCA revealed that the clustering was still present unfortunately some populations were so close to each other that it was hard to distinguish between them. A look at an inset at a higher resolution were the clustering occurs (Figure 21) reveals that the closest populations to the Maltese one, apart from Palermo which remained the closest, were now Sardinia, South Italy, Spain, Libya and Lebanon.

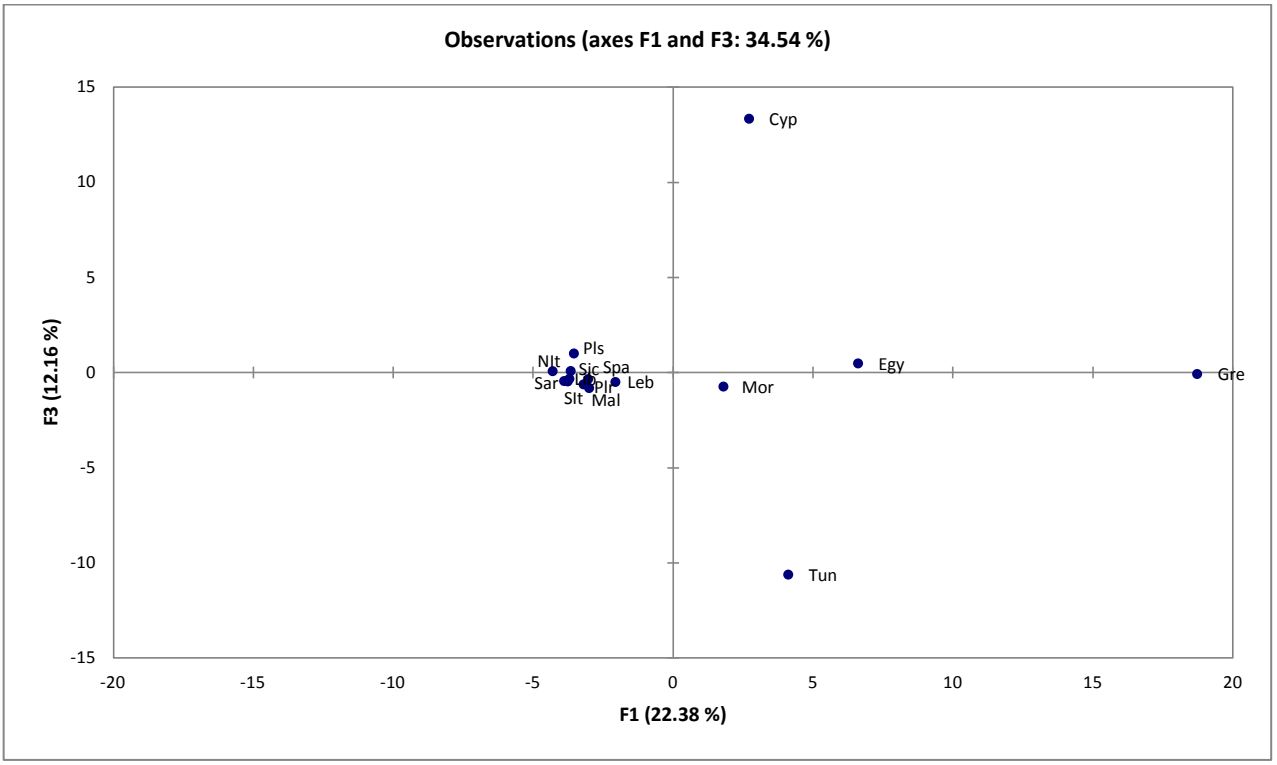

Figure 20: PC 1 vs PC 3 of individual polymorphism analysis.

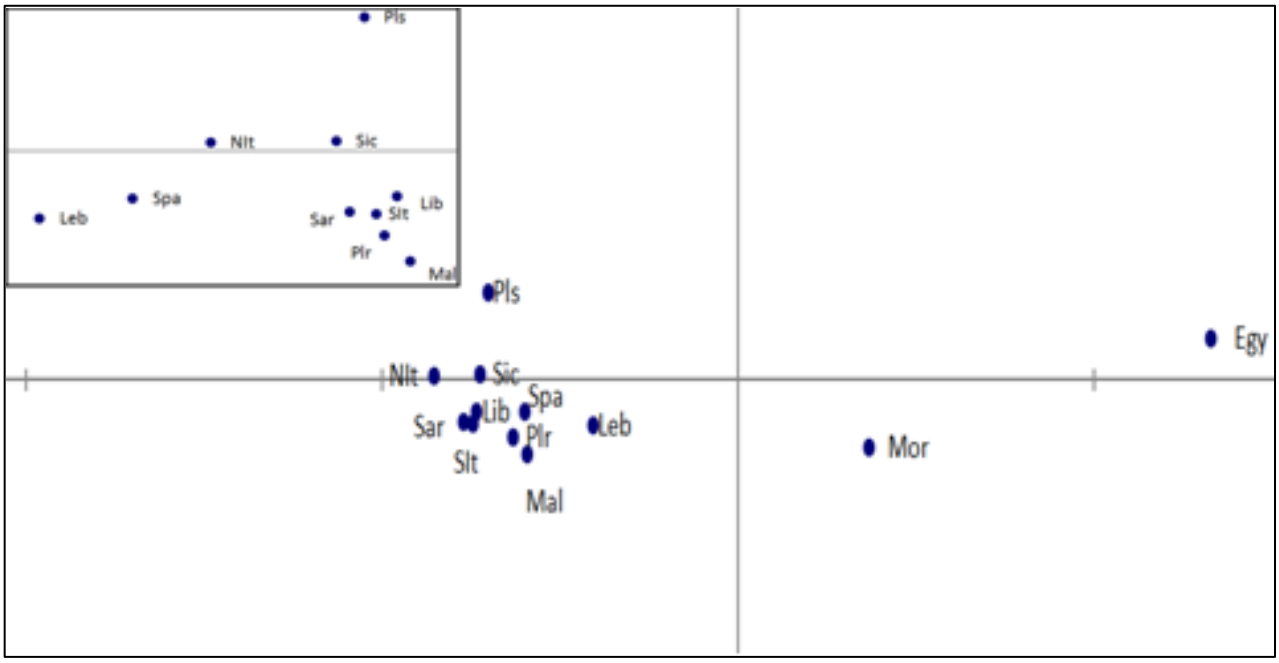

Figure 21: Inset of PC 1 vs PC 3 of polymorphism analysis.

The last analysis was conducted on PC 2 vs PC 3, which has the lowest variation of the plots at 26.44% (Figure 22). This analysis showed that Malta was once again close to Palermo.

Thus Palermo was the closest population to the Maltese one in all three plots. Following Palermo as the closest to the Maltese population, when the inset is observed (Figure 23) South Italy, Libya and Sardinia then follow as the closest populations. Thus these same populations cluster the closest to the Maltese population in all three plots, whilst other populations such as Palestine and Spain, come close to the Maltese population in some PC plots but distance themselves in others.

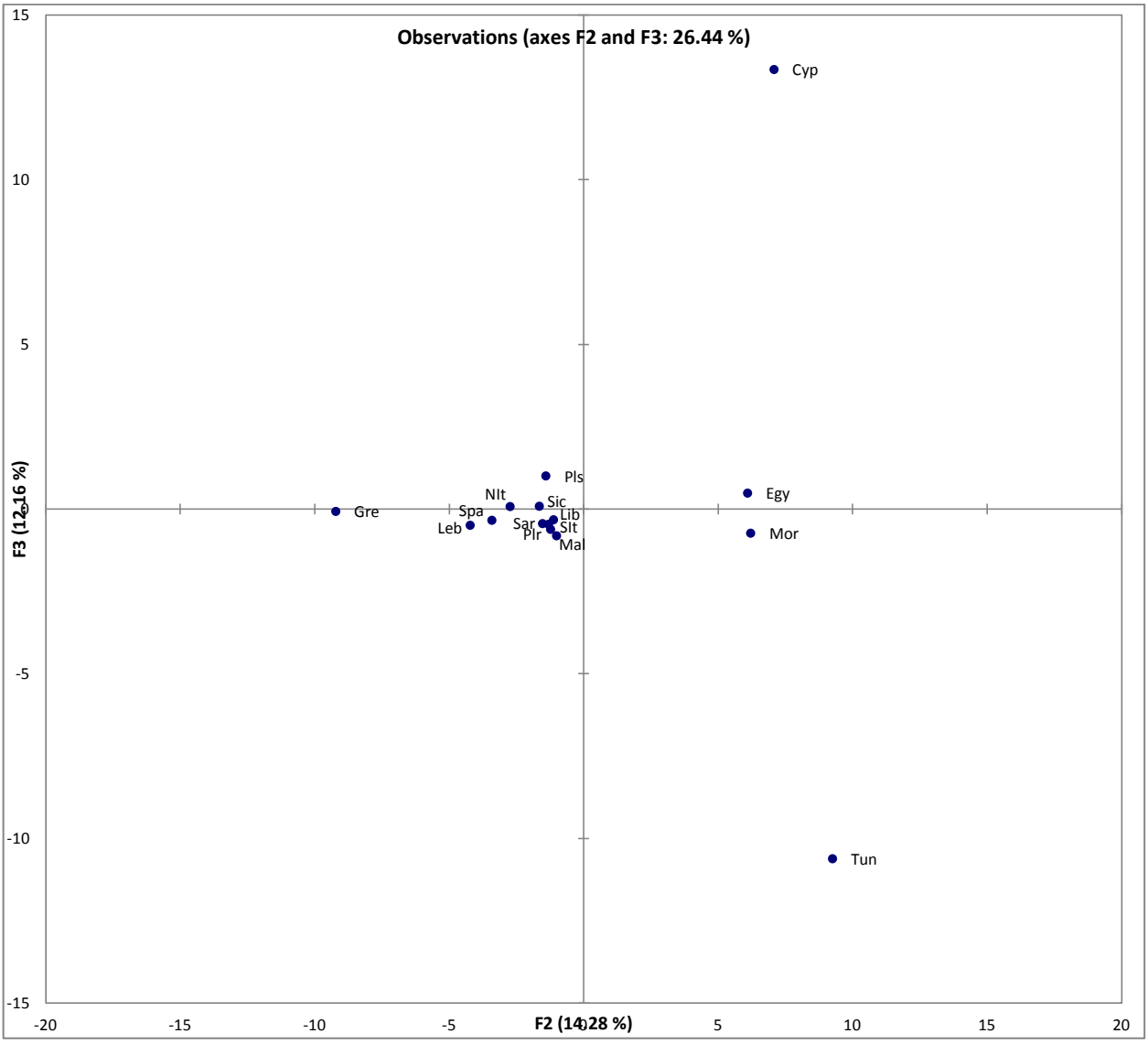

Figure 22: PCA of PC 3 vs PC 2.

Figure 23: Inset of above PCA showing the positions which cluster closest to the Maltese one.

4.5: Reduced Median Network Analysis

Haplogroup H which is the most common haplogroup in Europe was the first to be analysed. The complete reduced median network is presented in Appendix 7 as it is too large to be observed accurately. Thus a simplified version showing all the polymorphisms with a frequency of at least two is presented in Figure 24. The complete reduced median network showed the Maltese population clustering with Sicily and North Italy repeatedly. Sicily is also the only population with which the Maltese clustered exclusively. This happens three times. Maltese samples are also found very often as part of a cluster together with Greek samples (four times) and with North Italian samples (six times) An interesting fact is that the Maltese sample also clusters often with the Lebanese population (four times). Whilst the large population samples (Table 6) can be an explanation for the clustering of the Maltese samples with the Greek, Italian and Lebanese populations, this explanation is not suitable in

order to explain the amount of times the Maltese samples cluster with the Sicilian samples, especially the three times were the populations cluster exclusively.

The Maltese population forms the largest sample in a cluster with Libyan and North Italian samples. This is interesting since the other two populations have got a larger sample size then the Maltese one.

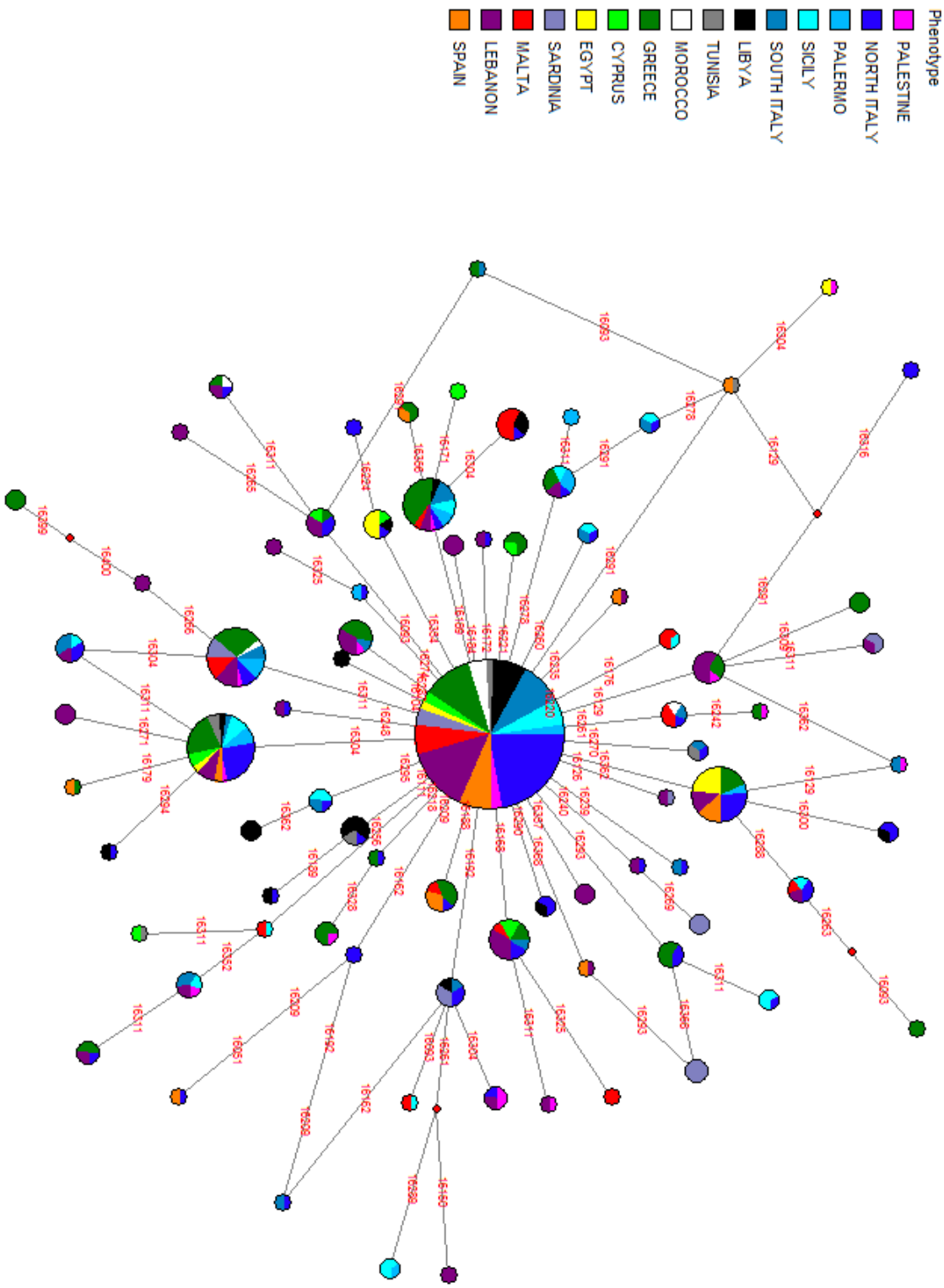

Figure 24: Haplogroup H.

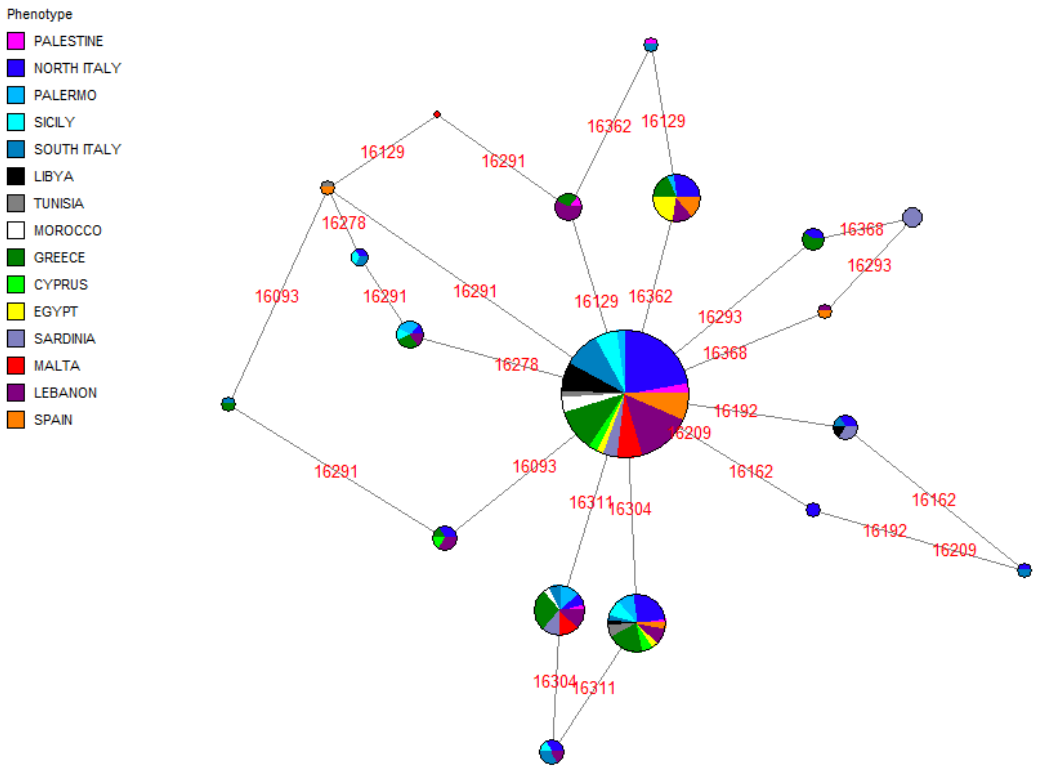

Figure 25: Haplogroup H torso.

The Maltese samples situated in the torso (Figure 25) shows similarity to Richards *et al.* (1998) which shows a cluster of Mediterranean-West (Iberia) at mutation 16311 from the rCRS. Whilst the Spanish population is not part of this cluster in this study, a cluster from a large area of the Mediterranean including Malta forms at this node. This can be explained due to the larger dataset used in this study. Maltese samples in the torso are also found carrying the rCRS, as expected.

Haplogroup J (Figure 26) shows the Maltese population clustering most often with South Italian samples. This happens two times. One of these times is in a cluster with Sicily and Lebanon, whilst the other time sees the Maltese population cluster with North Italian and Greek samples.

The fact that Sicily clusters once with Malta is interesting especially when this is compared to the other populations which cluster with Malta such as North Italy, Greece and Lebanon which all have at least double the sample size of the Sicilian one (154 samples).

The Maltese samples which cluster after mutations 16168-16264-16295 are the haplotypes which have been classified as forming haplogroup JT. These again show, as has been seen in the Maltese reduced median network cluster, that they are not part of haplogroup J. Of interest is that the other Maltese cluster is situated two mutations away from a South Italian cluster, which also yields a Sicilian sample.

The Maltese samples in the torso (Figure 27) of haplogroup J are found in conjunction with Sicily, South Italy and Lebanon. The polymorphism is also found in Richards *et al.* (1998) and is made up from British samples in that study. This means that the cluster can be found in both the Mediterranean and also in Northern Europe.

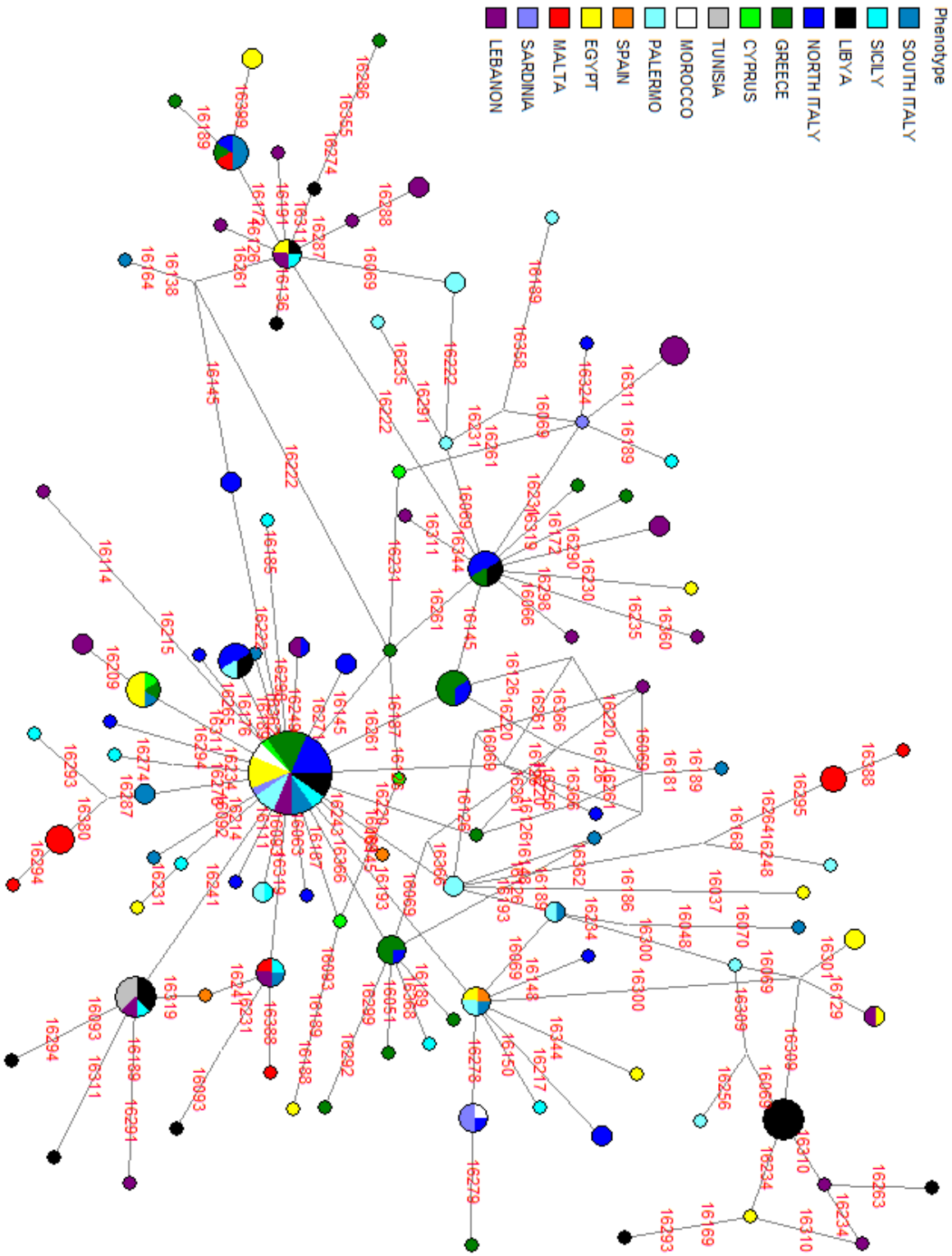

Figure 26: Haplogroup J.

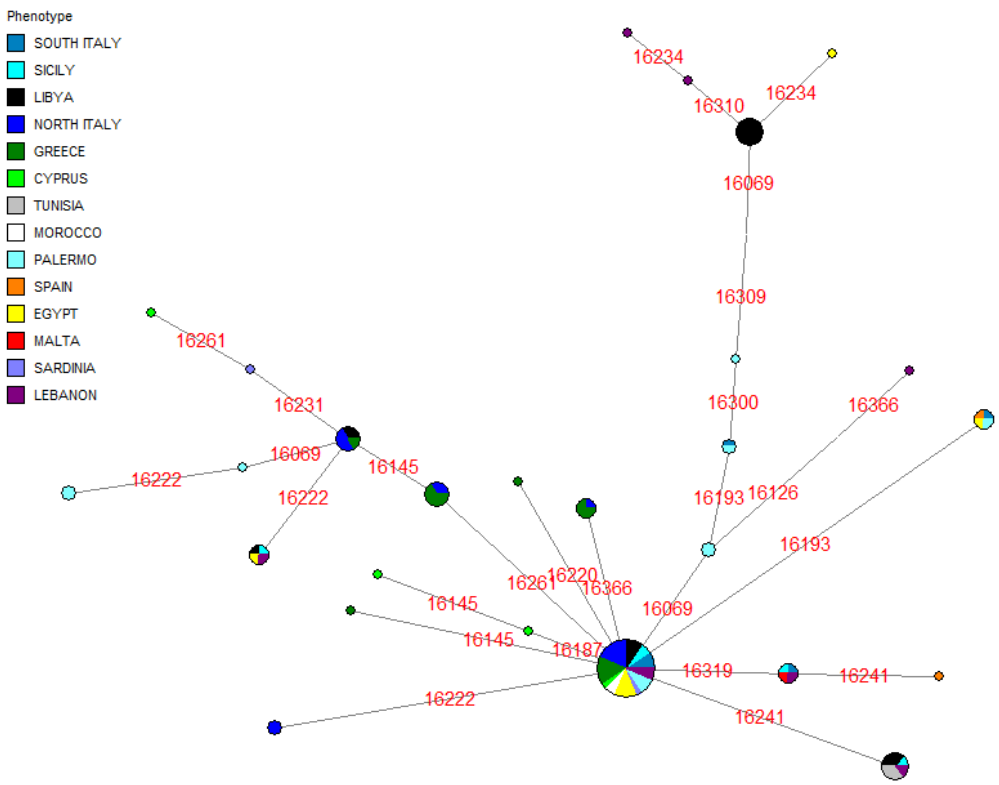

Figure 27: Torso of Haplogroup J.

In the haplogroup T reduced median network (Figure 28) the Maltese population is seen several times clustering with different populations from around the Mediterranean. The most common populations the Maltese population clusters to are the North Italian and the Sicilian population at five and four times respectively. Of interest in this network is the Maltese cluster with mutation 16147 which emerges from an Egyptian sample and leads to another Egyptian cluster of two samples. This Maltese cluster of two samples also leads to two other Maltese samples.

A Maltese cluster of two samples is one mutation away and emerges from a cluster containing Egyptian, North Italian and Palestinian samples. However it must be noted that the Egyptian sample is almost double that of Sicily and only clusters with a Maltese sample

one time when compared to the four times the Sicilian sample clusters. Greece and Spain also cluster with the Maltese samples both at three times each.

A view of the torso of haplogroup T (Figure 29) shows that the mutations 16163, 16186 and 16189 which separate T from T1 are also present in the Mediterranean samples, with a Greek sample acting as a vector between 16163 and 16189. The mutations 16296 and 16304 which reflect haplogroup T2 are also reflected in the Torso with the Maltese sample being present at both mutations.

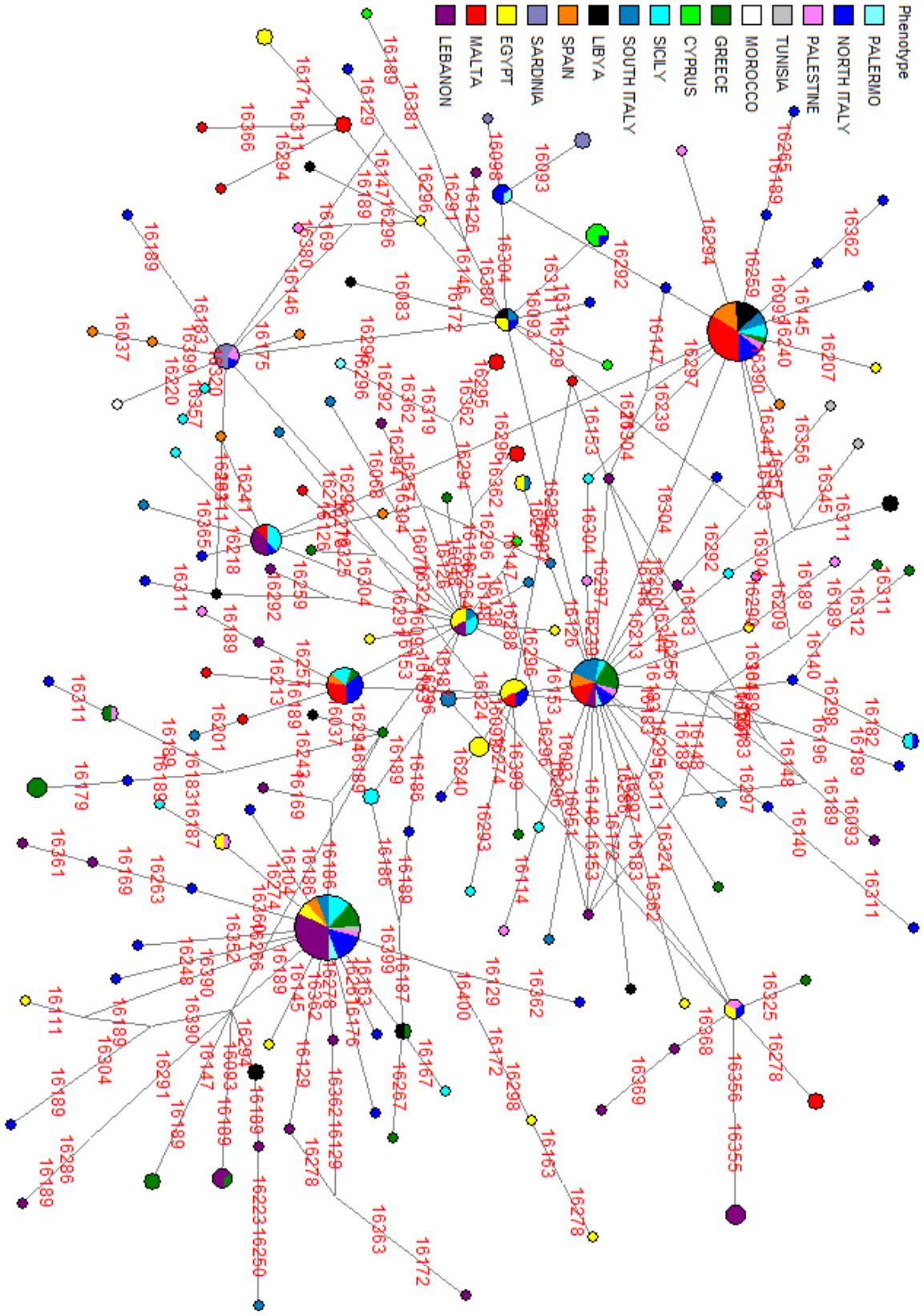

Figure 28: Haplogroup T.

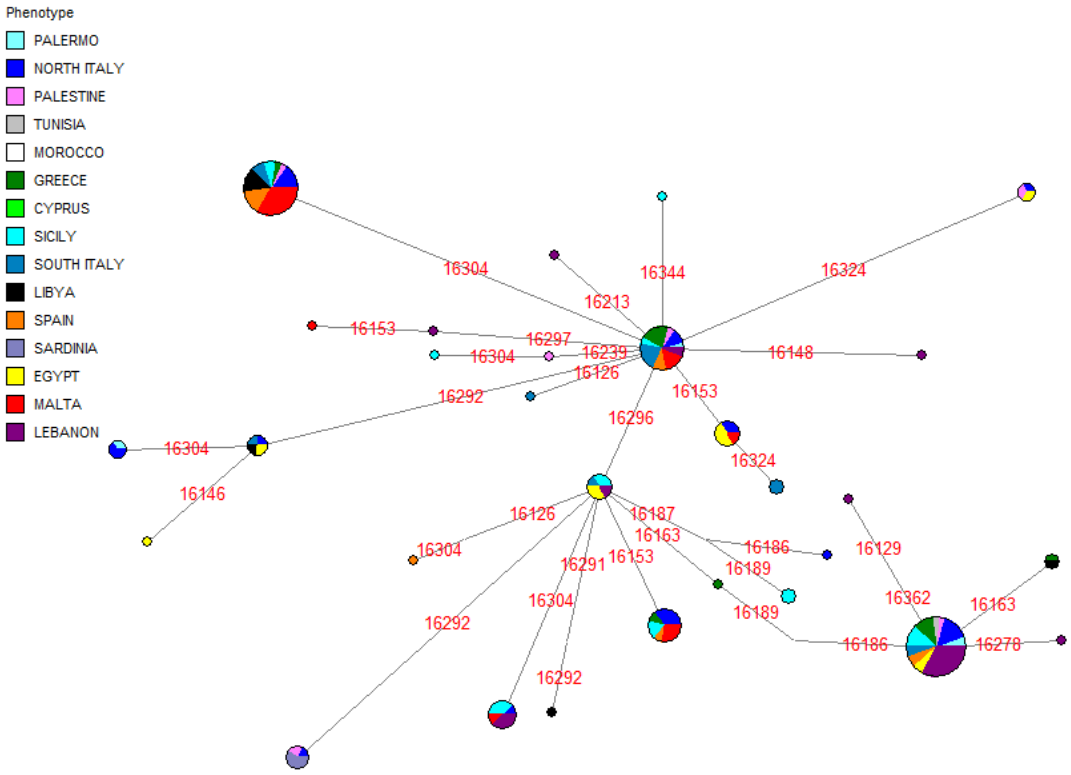

Figure 29: Haplogroup T torso.

The network of haplogroup L is found in Appendix 7. Due to the large number of nodes which are present a simplified version made of polymorphisms which appear at least twice in the dataset is presented in Figure 30. In order to make sure that all the Maltese samples pertaining to this haplogroup are shown, each one of them was given a frequency of two. Haplogroup L is very interesting due to the Maltese samples almost always clustering with other Maltese samples only. Of special interest is the only time that the Maltese samples cluster with another population. This population is Cyprus and not a North African population as might be expected due to the L haplogroups being African, and they being found at very low levels in Europe.

The node where the Maltese and Cypriot samples form a cluster emerges from a cluster made of Tunisian and Spanish samples and is identified as haplogroup L2.

The large amount of Maltese samples with this haplotype indicates either that a population bottleneck occurred in the Maltese population, or that the polymorphism has been present for a long time. Mutations of the offshoots from this cluster were checked multiple times in order to be sure that they were not sequencing errors.

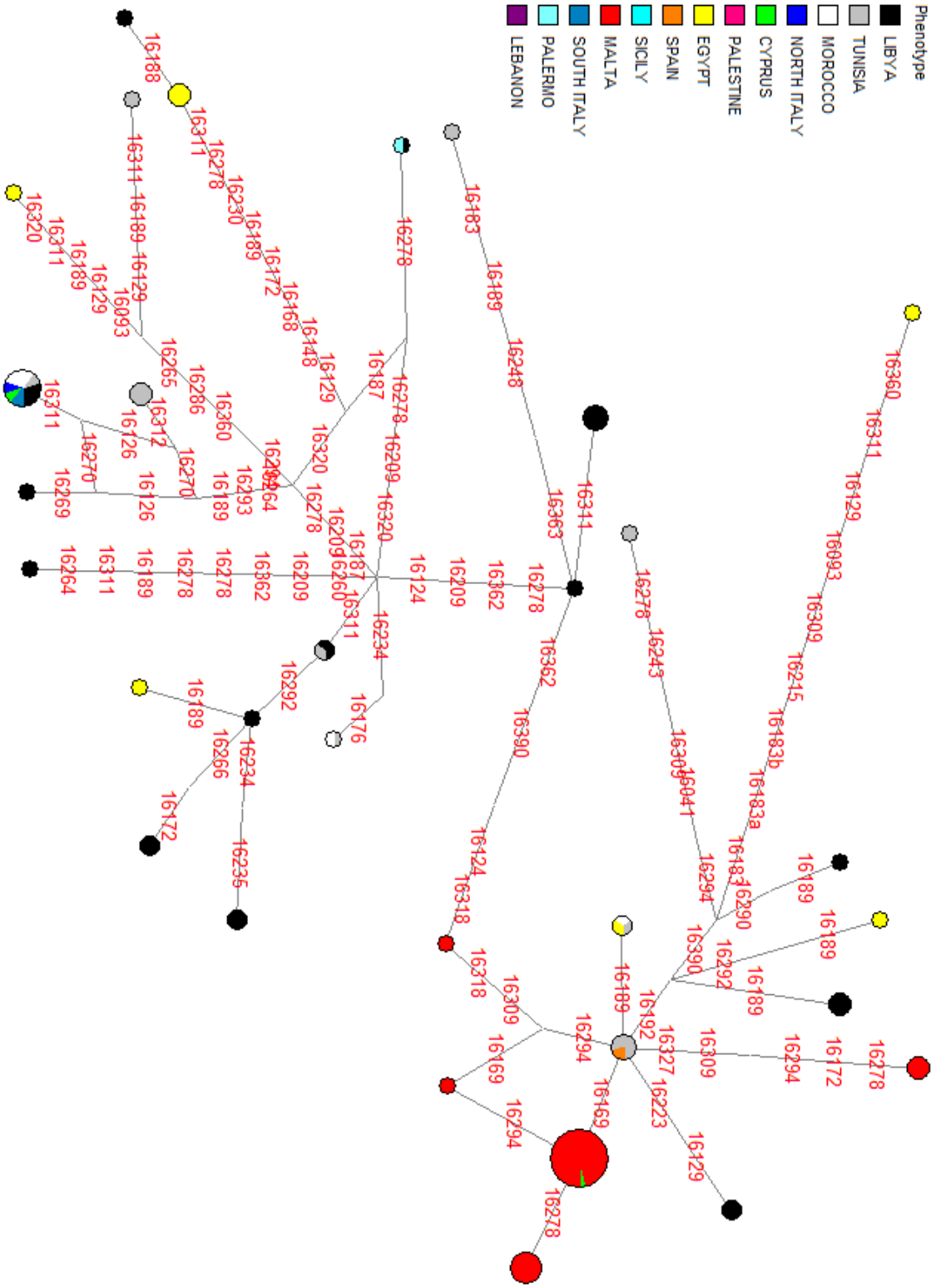

Figure 30: Haplogroup L reduced median network showing haplotypes at a frequency of 2 or more. Except for the Maltese population which is fully represented in the network.

Haplogroup U (Figure 31) shows the Maltese samples cluster in two nodes. Both nodes are made up exclusively of Maltese samples, but whilst one of the nodes is not very informative due to it being a large number of mutations away from the closest cluster it might emerge from, the other node emerges from a cluster made of Palestinian, Lebanese and Cypriot samples. This indicates a Near Eastern influence on the Maltese population. Both nodes are represented in the torso of the haplogroup (Figure 32).

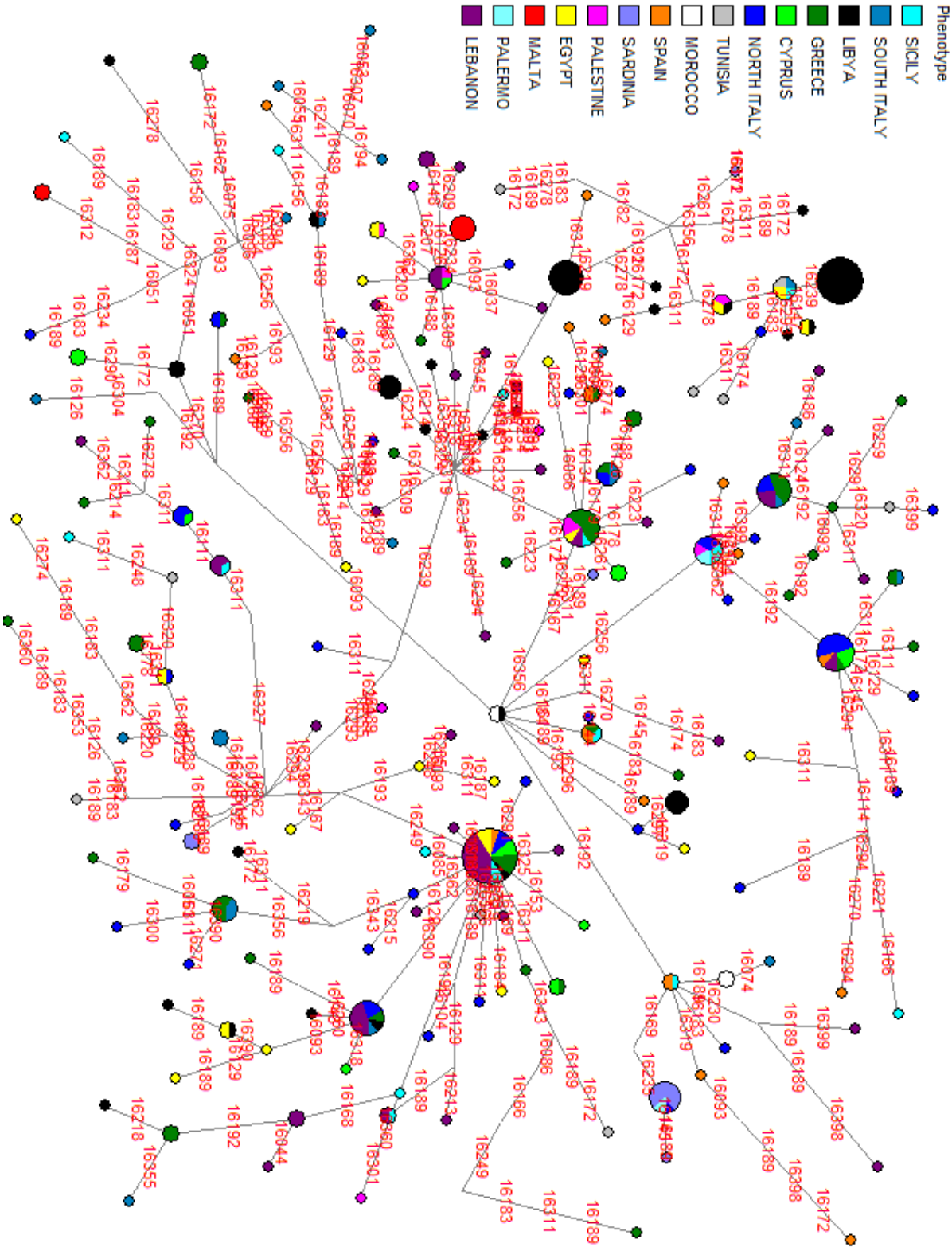

Figure 31: Haplogroup U.

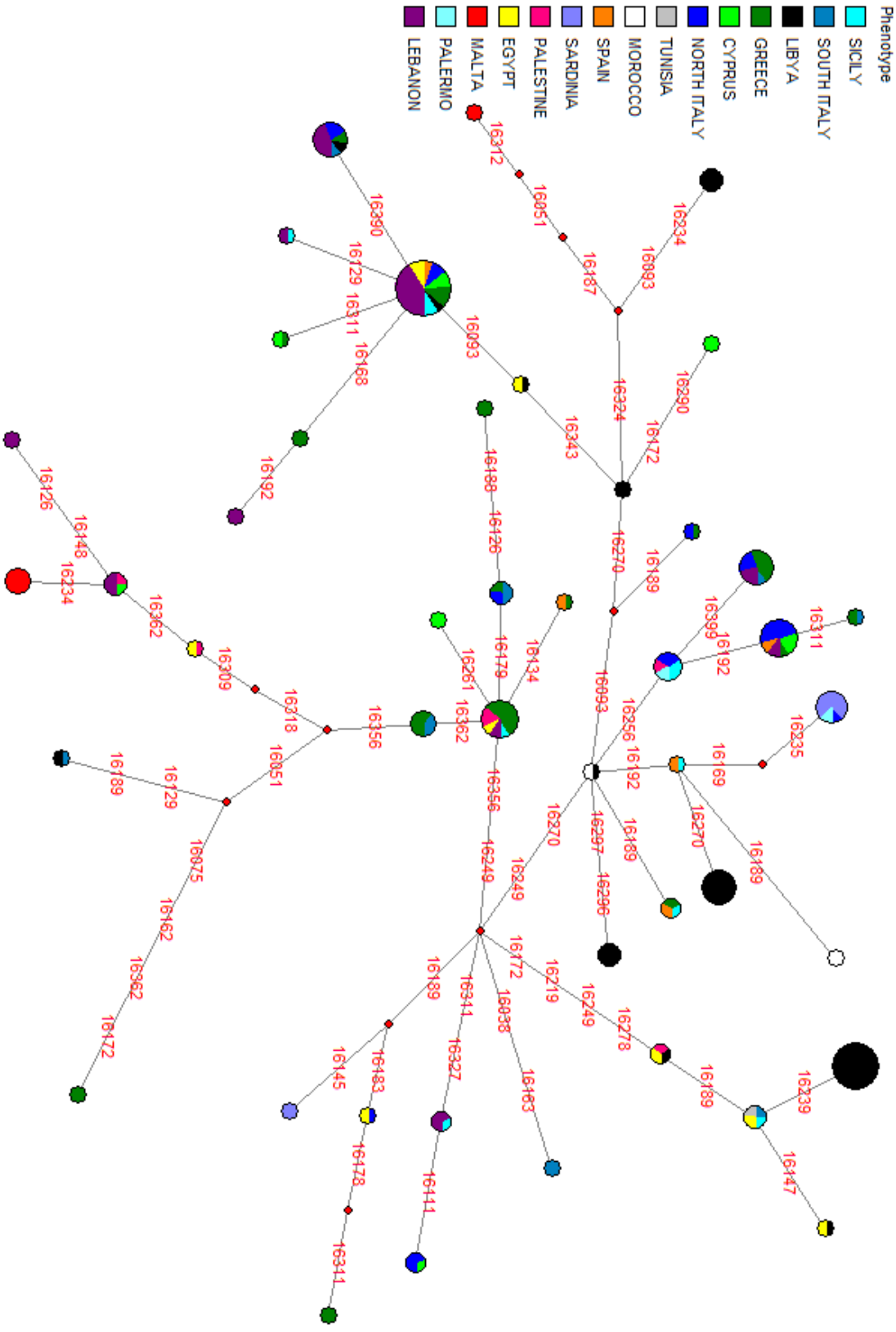

Figure 32: Haplogroup U showing haplotypes found more than one time. Both clusters containing Maltese samples are represented.

The reduced median network (Figure 33) shows haplogroups HV and V and their variants with a Maltese sample clustering in a node which contains samples from all the populations, this thus makes it difficult to garner any information from it. Of interest is that the cluster yields another two Maltese samples. Analyzing these ancestral haplogroups is always difficult without the information from coding regions about the specific samples.

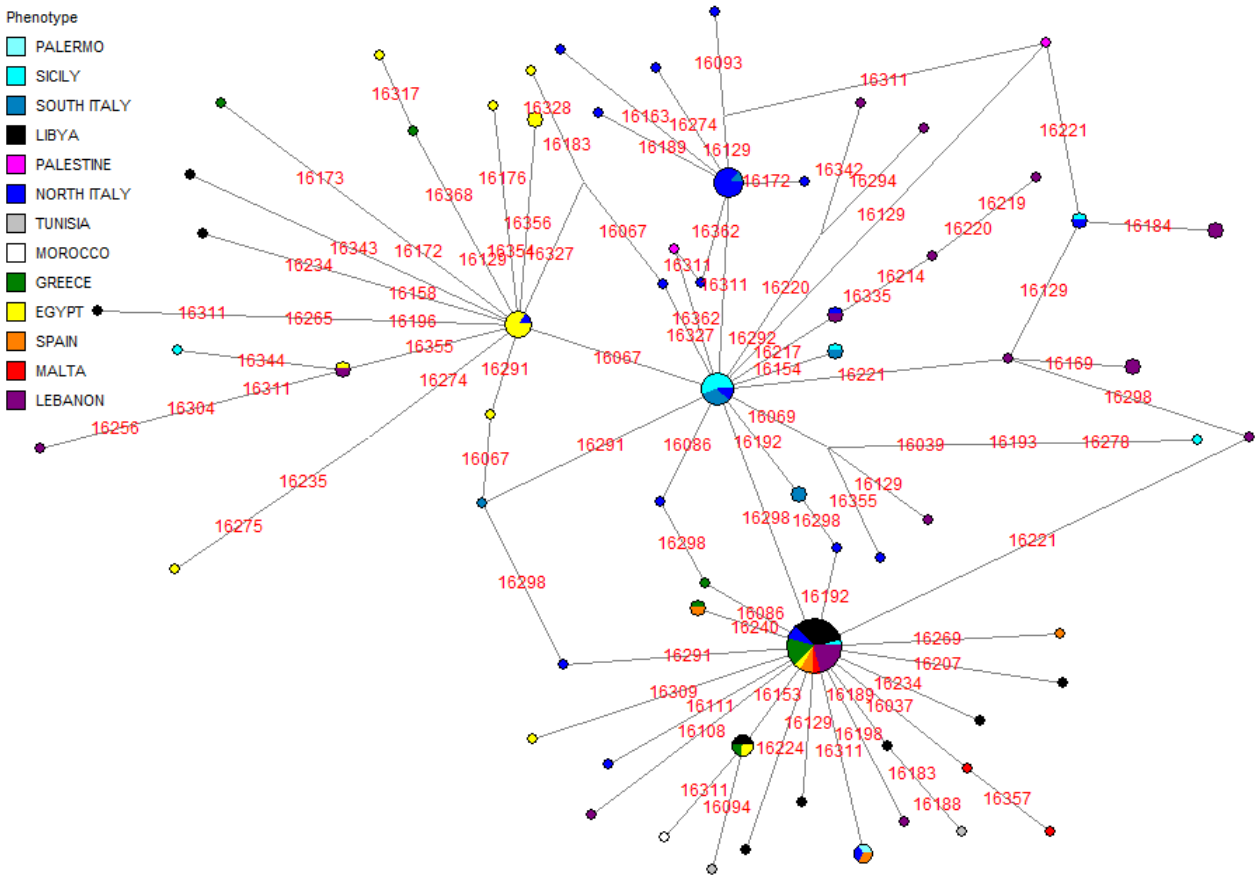

Figure 33: Haplogroups HV and V.

Haplogroup W (Figure 34) shows that the only Maltese sample clusters together with a Palestinian and a Greek sample. This is indicative that the Maltese population must have acquired this specific W haplotype from the Eastern Mediterranean, especially when one considers that Italian samples make up almost half of the haplotypes being represented here.

The lack of any Lebanese samples clustering together with this cluster especially due to the large sampling size indicated that the haplotype possibly came to Malta through a Greek pathway since there are some commonalities between Greek and Maltese history, and if the haplotype had come to Malta from Greece through Italy an Italian sample would have been expected to be observed in the cluster. The same commonality in history cannot be said to exist between Malta and Palestine.

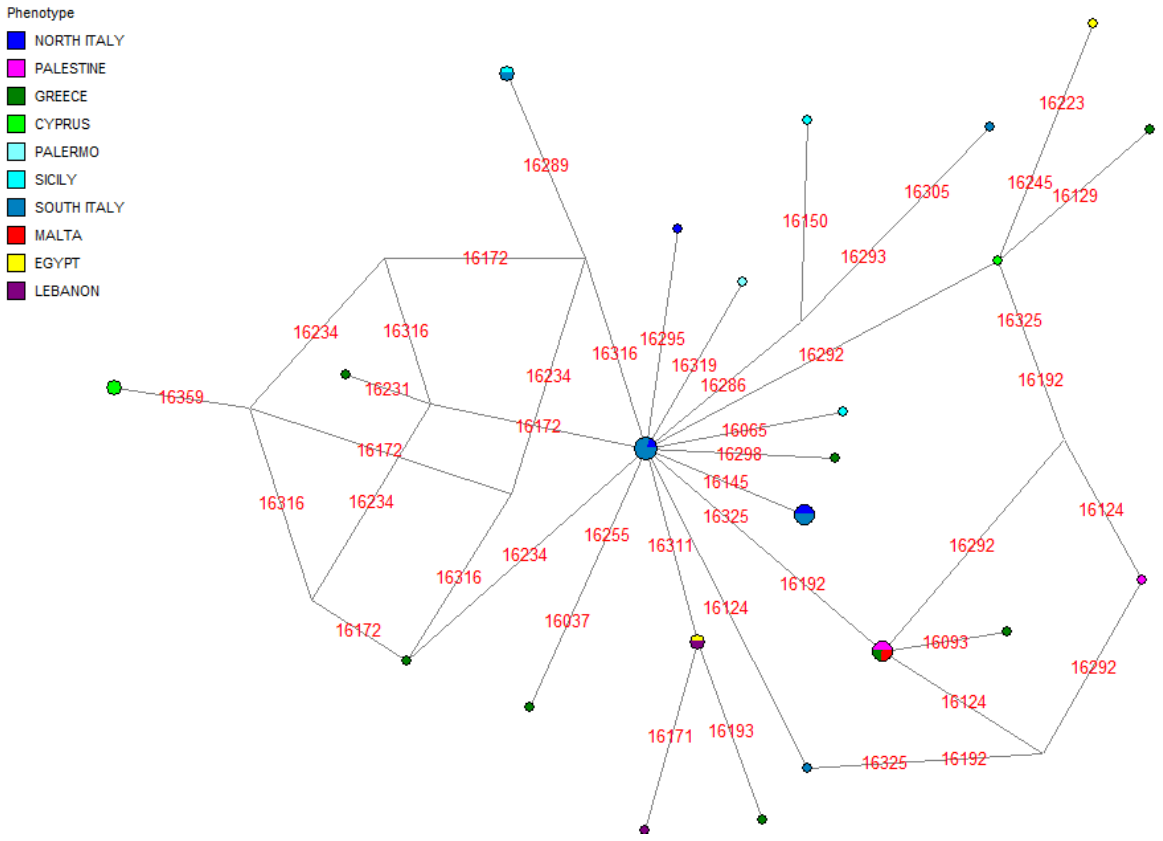

Figure 34: Haplogroup W.

The Maltese samples in haplogroup K (Figure 35) are all part of just two clusters. One of the clusters is shared exclusively by Maltese samples and is found two mutations away from the main cluster, whilst the other samples form part of cluster K1 which contains samples from almost all the populations in the study, and thus is not very informative. The extensive geographical spread of K1 can be viewed in the torso (Figure 36) were in Richards *et al.*

(1998) mutation 16093 results in samples from the North-West, North-Central, Alpine, Mediterranean-Central and Mediterranean-West clustering together, and in this study almost all the populations are represented in the cluster.

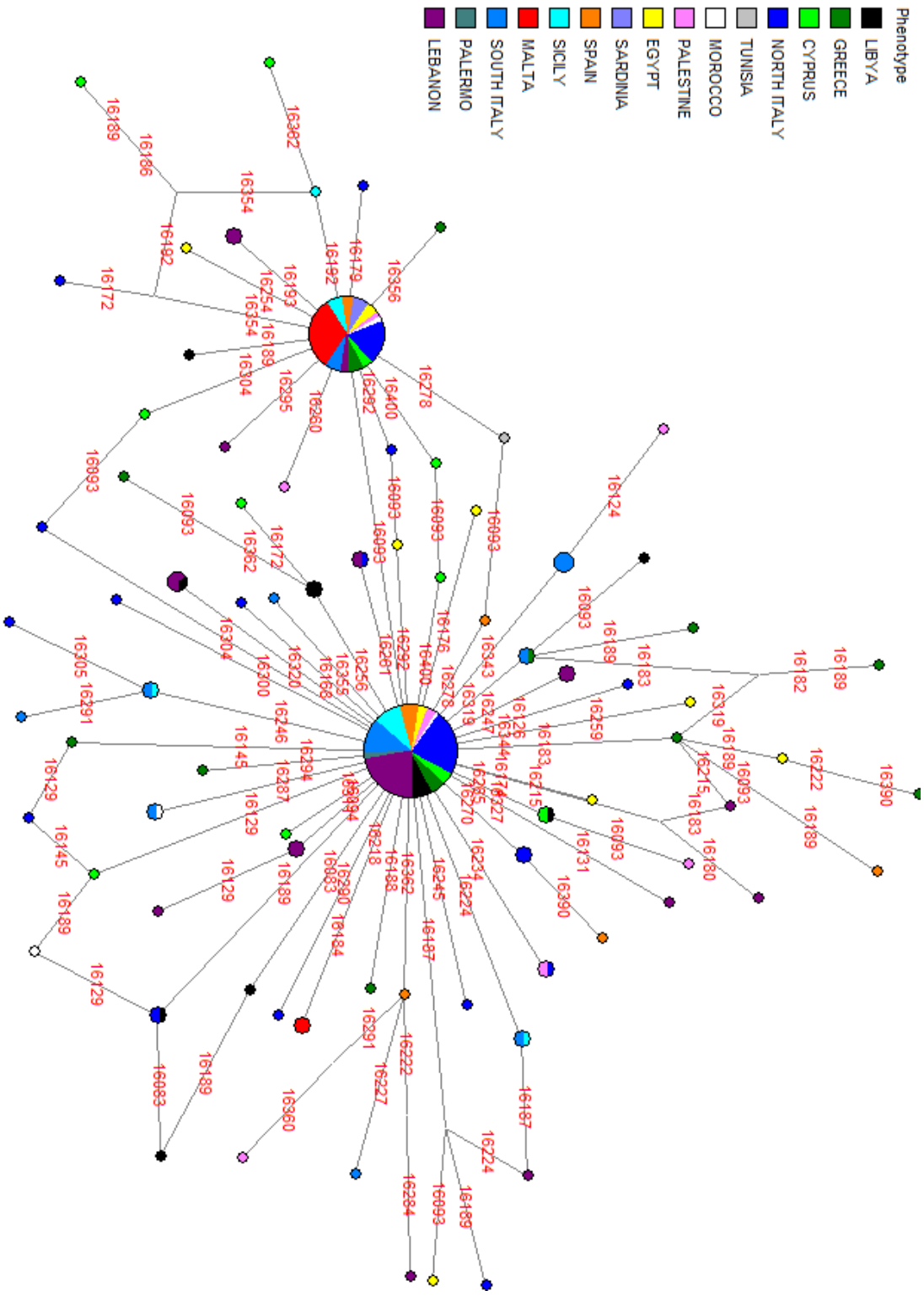

Figure 35: Haplogroup K.

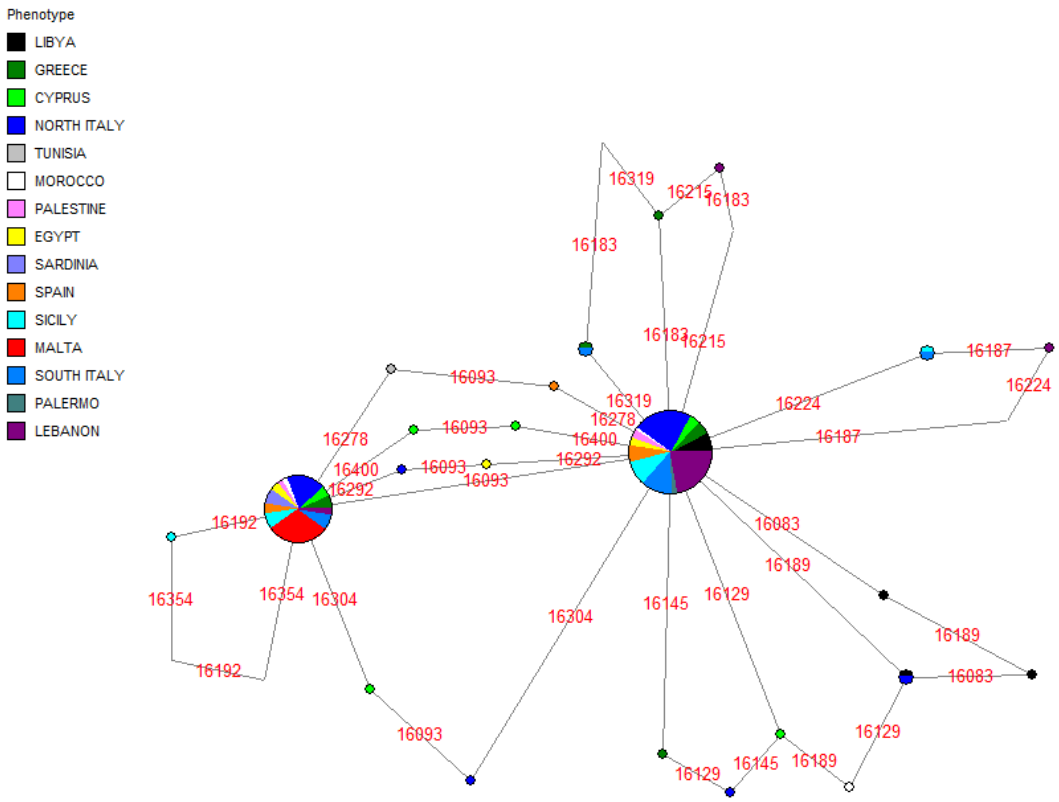

Figure 36: Torso of haplogroup K.

In haplogroup X (Figure 37) the only Maltese sample clusters with a Greek sample. It is interesting that a haplogroup rare in itself in the Maltese population clusters with a population which is not as close to the Maltese one geographically as Italy or Tunisia is. This relationship is strengthened when the amount of times in other haplogroups where the Maltese and the Greek samples share a haplotype is considered.

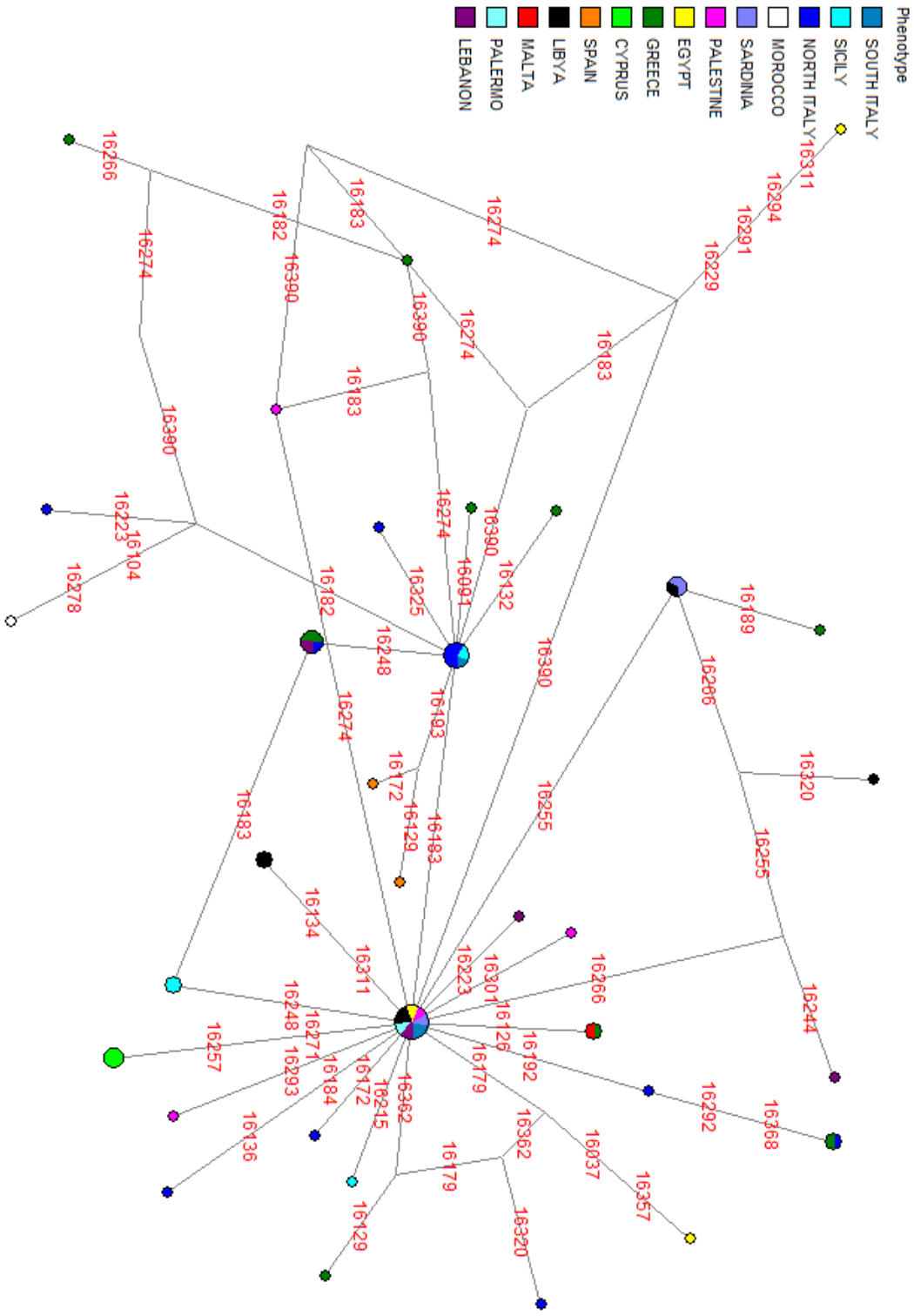

Figure 37: Reduced median network of haplogroup X.

4.6: Dating of Clusters

An attempt to date clusters was made on some of the Maltese samples, where these samples clustered together exclusively in an attempt to be able to determine during which period did these mutations arise in the Maltese population. The mutation rate used was 1 substitution every 16677 years between positions 16051 and 16400, with the exclusion of positions 16182C, 16183C and 16194C according to the calculator devised by Soares *et al.* (2009). Unfortunately the results obtained were not informative due to the given age and range of the clusters spanning a Maltese prehistory and history.

Chapter 5: Y-STR analysis of the Maltese population in relation to other populations from around the Mediterranean

5.1: STR data collection from the Maltese population

Y-STR analysis was conducted in order to determine the similarities and differences the Maltese population's male lineage has with other populations from around the Mediterranean. The reason for this study is that many rulers of Malta throughout the ages ruled the islands using very few men who came alone and did not bring their families with them, such as the Knights of Saint John. Thus this study was an attempt to see if a certain period in history can be identified when the males who moved into the island left their offspring there as well.

The main problem this analysis had was that due to the dye attachment method being developed in house, its success rate was not as high as that of forensic kits targeting the same STRs. This meant that the analysis involved some missing data from the Maltese population, where STR results were not obtained after two attempts at genotyping and thus were extrapolated (Appendix 8). These can be identified in the dataset as they are represented by the number 0 instead of the number of repeats in the full STR dataset (Appendix 9). This issue meant that some STRs which were analysed had to be discarded from the final analysis due to a low success rate.

A problem which was also encountered in the analysis arose from the large number of studies conducted on different populations in the Mediterranean region which had used a

variety of different Y-STR markers. Ony five of the Y-STRs typed for the Maltese population in this study could be compared with a large amount of other populations in the Mediterranean as only these five were found in all of the other studies. To use a larger amount of the Y-STR dataset would have meant having to exclude some of the comparison populations for which these additional STRs had not been typed, and this was done in the the eight Y-STR analysis. STR analysis was conducted by the methods described in the Materials and Methods chapter. The results were then compared with other populations from around the Mediterranean region (Table 7).

Population (Key)	Sample Size (n)	Reference
Greece (GRE)	199	Kovatsi (*et al*. 2009)
Italy (ITA)	292	Brisighelli (*et al.* 2012)
Lebanon (LEB)	587	Zalloua (*et al.* 2008)
Libya (LIB)	239	Elmrghni (*et al.* 2012)
Malta (MAL)	142	Present Study
Modena (MOD)	100	Rossi (*et al.* 1999)
Sardinia (SAR)	69	Ghiani & Vona (2002)
Sicily (SIC)	255	Robino (*et al.* 2006)
Spain (SPA)	148	Martin (*et al*. 2004)
Tunisia (TUN)	244	Cherni (*et al.* 2005)
Turkey (TUR)	208	Rűstemoglu (et al. 2011)

Table 7: Comparative populations used for the Maltese Y-STR study.

5.2: STR data validation

As a proof of concept, in order to to make sure that the variance obtained had not been random, the ratio of the observed variability against the estimated variance was measured over 99 permutations (Figure 38).

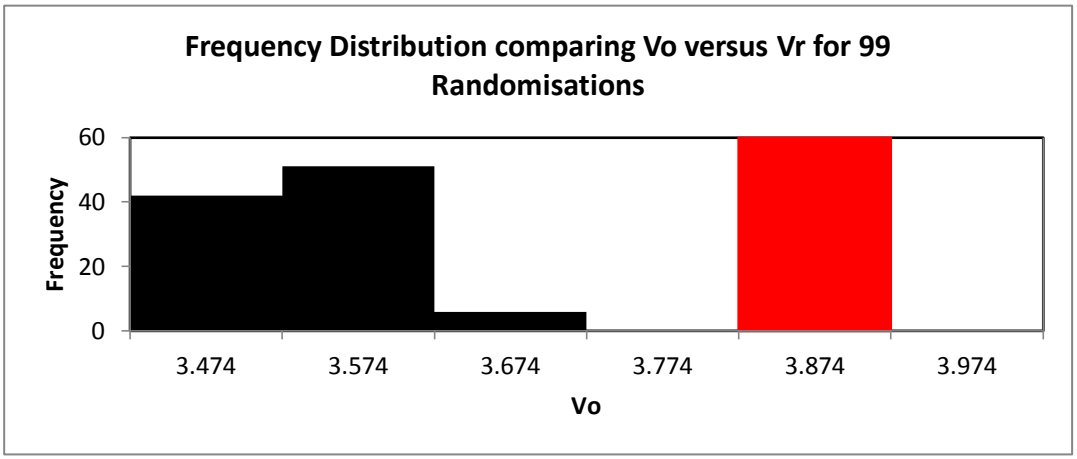

Figure 38: Bargraph showing the difference between the observed variance (red) when compared to the estimated variance calculated using GenAlEx. For each sample *Vr* is calculated as the variance of the randomized data set, and the probability of observing a Vr value as extreme as the observed value (Vo) is calculated (Peakall and Smouse 2010).

Figure 38 shows the difference between the range of expected extreme variance over 99 permutations of the dataset, and the variance which was obtained. This means that the variance in the STR dataset was not random but linked to different population lines.

5.3: Analysis of Y-chromosomal STR variance amongst populations from around the Mediterranean

In the mtDNA analysis, low molecular variance between different populations was observed when compared to the variance within the populations. In order to see if this pattern also repeats itself in the male side of the study, an AMOVA was conducted with the STRs that were going to be included in the inter-population study using both GenAlEx (Figure 39) and Arlequin (Figure 40).

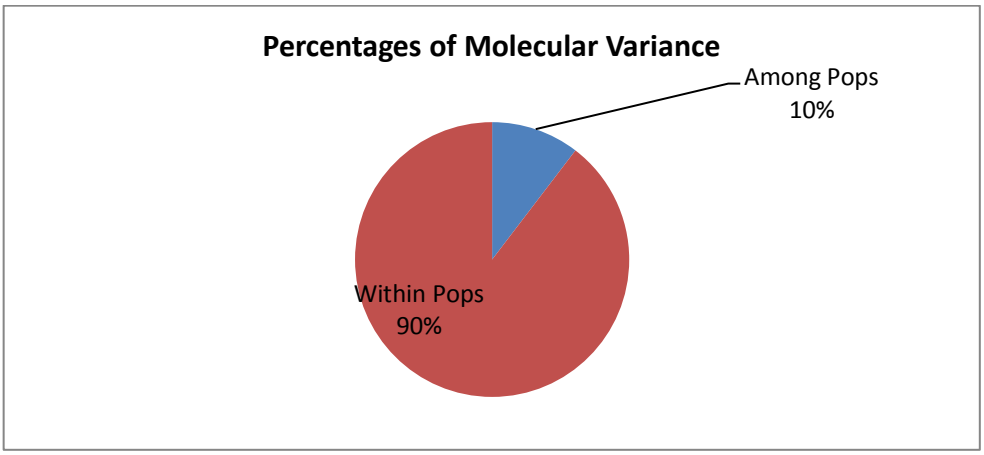

Figure 39: AMOVA showing amongst and within population variance over 999 permutations calculated using GenAlEx.

```
=====================================================
Global AMOVA results as a weighted average over loci
=====================================================

-----------------------------------------------------
AMOVA design and results (average over 8 loci):
-----------------------------------------------------

------------------------------------------------------------------------
 Source of              Sum of          Variance           Percentage
 variation              squares         components         variation
------------------------------------------------------------------------
 Among
 populations           1168.855            0.75321           10.14951

 Within
 populations          12216.966            6.66793           89.85049
------------------------------------------------------------------------
 Total                13385.821            7.42114
------------------------------------------------------------------------


Average F-Statistics over all loci
 Fixation Indices
       FST :      0.1015
```

Figure 40: AMOVA calculated using Arlequin.

The result showed an estimated variance of around 10% between populations, whilst the other 90% of variance was found amongst the populations in 999 permutations in both programs.

This result indicates that the populations being studied display a greater diversity between populations in their Y chromosome than in their mtDNA, since the AMOVA result for the eight STR loci in all the samples showed a molecular variance of 10% in the populations when compared to a variance of 2% for the mtDNA data.

5.4: Principle component analysis of Mediterranean STR data

A pairwise distance matrix (Table 8) of the variance between populations was constructed in order to determine the Y-STR genetic distance between different populations in the Mediterranean region using 5 Y-STRs. This distance matrix showed that the Maltese population was genetically closest to the Sicilian and Sardinian populations.

GRE	ITA	LEB	LIB	MAL	MOD	SAR	SIC	SPA	TUN	TUR	
0.000											GRE
0.118	0.000										ITA
0.056	0.131	0.000									LEB
0.074	0.157	0.075	0.000								LIB
0.037	0.097	0.057	0.081	0.000							MAL
0.169	0.017	0.196	0.221	0.115	0.000						MOD
0.014	0.104	0.048	0.103	0.027	0.147	0.000					SAR
0.025	0.053	0.034	0.073	0.027	0.082	0.013	0.000				SIC
0.144	0.010	0.170	0.190	0.101	0.000	0.121	0.066	0.000			SPA
0.105	0.249	0.222	0.094	0.157	0.335	0.174	0.145	0.297	0.000		TUN
0.113	0.178	0.145	0.132	0.079	0.147	0.115	0.102	0.134	0.273	0.000	TUR

Table 8: Y-STR genetic distance calculated from 5 Y-STRs between 11 populations from around the Mediterranean.

This distance matrix was also generated with the 8 Y-STR dataset of the seven populations which had the required Y-STRs in their studies (Table 9). The results showed Malta being closest to the Lebanese and Italian samples, followed by Greece and Spain.

Lebanon	Greece	Libya	Malta	Spain	Tunisia	Italy	
0.000							Lebanon
0.028	0.000						Greece
0.058	0.075	0.000					Libya
0.053	0.072	0.116	0.000				Malta
0.068	0.099	0.160	0.082	0.000			Spain
0.128	0.099	0.078	0.167	0.203	0.000		Tunisia
0.026	0.031	0.100	0.054	0.040	0.157	0.000	Italy

Table 9: Y-STR genetic distance calculated from 8 Y-STRs of 7 populations from around the Mediterranean.

A PCA was then conducted in order to determine the relationship between the populations in a multi-dimensional manner. This PCA was conducted using GenAlEx 6.41 as this program enables individual samples to be grouped into their respective populations. The details of the PCA can be found in Appendix 10.

The first principal component plot (Figure 41) places the Italian and the Sicilian populations as the closest to Malta. The closest populations after these two are the Sardinian and Turkish populations.

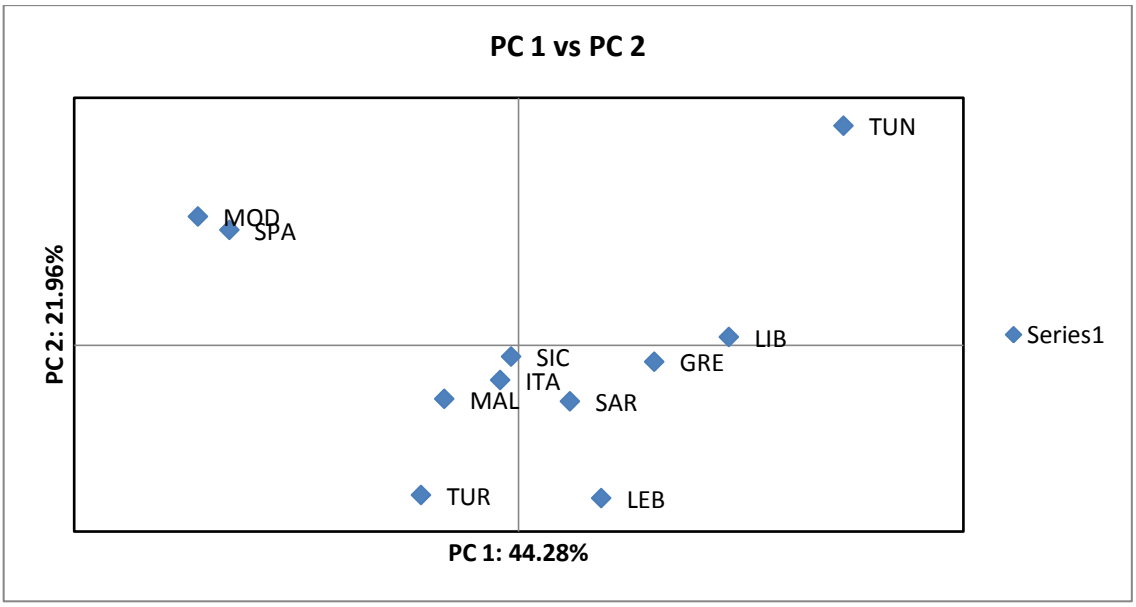

Figure 41: PC 1 vs PC 2 of 5 Y-STR. The population key is given in Table 7.

The second PC plot (Figure 42), which compares the 1st and 3rd PCs, also shows the same configuration, with Sicily and Italy being the closest to the Maltese population, and the Sardinian population following them. Another factor of interest is the Libyan and Tunisian populations, which whilst they are found in the same quarter of the plot as before, are closer together thus showing more pronounced relatedness.

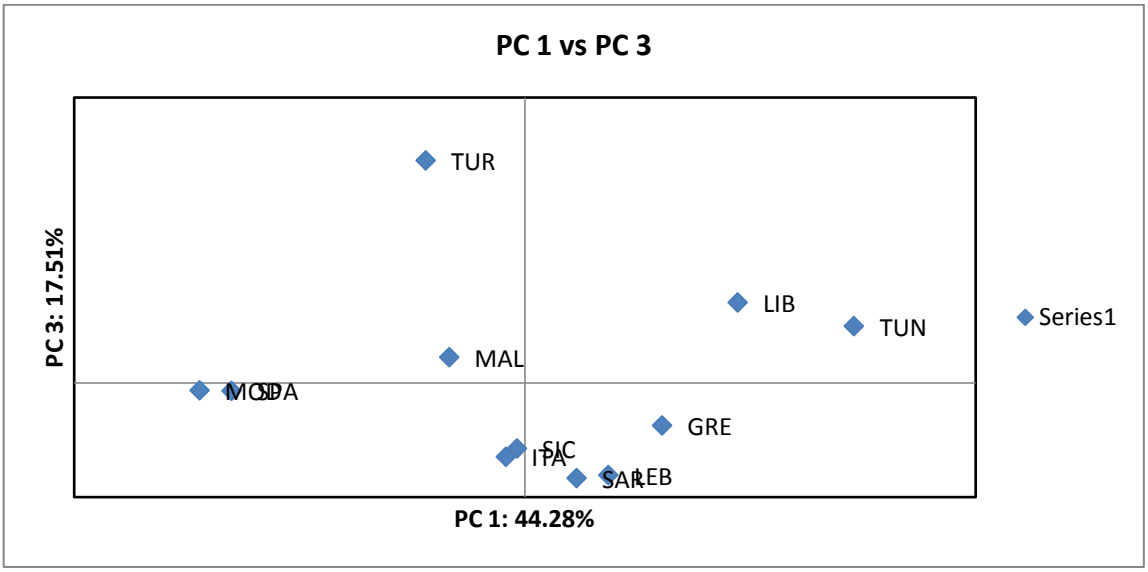

Figure 42: PC 1 vs PC 3 of 5 Y-STR.

The third plot, comparing the 2nd and the 3rd PCs, shows the Maltese population as being closest to the Greek one followed by the Sicilian, Italian and the Libyan ones (Figure 43). This plot additionally shows the Spanish sample and the sample from Modena clustering closely together. This plot, although it varies slightly from the previous two in terms of which population is the closest to Malta, still mantains the general trend of the other Western Mediterranean islands being found close to the Maltese one.

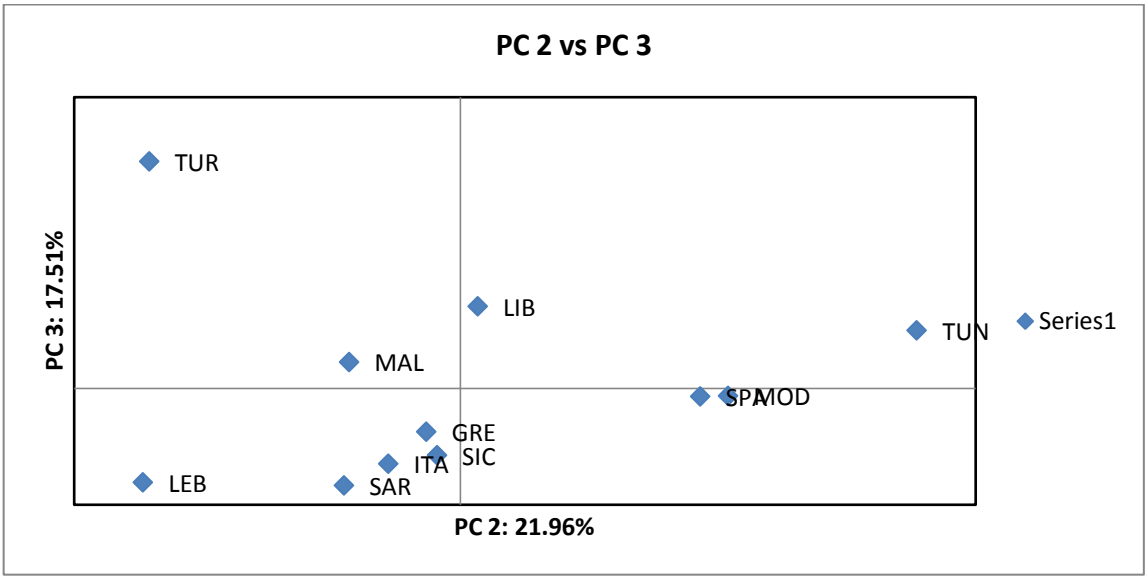

Figure 43: PC 2 vs PC 3 of 5 Y-STR.

The results of the Y chromosome STR study shown here agree with the results obtained by Y chromosomal haplogroup analysis conducted by Di Gaetano *et. al*. (2009) which placed Malta close to West Sicily, Calabria, Sardinia and East Sicily. This gives more strength to the present analysis, as it means that the STR data from Y chromosomes in this study agrees with the SNP data conducted on Y chromosome with a different sample of the Maltese population. Of interest is also that whilst Turkey is far removed from other populations in the Mediterranean region, its closest population in all of the three PC plots is Malta, even though Malta always has closer populations to it than the Turkish one. This might be a reflection of converted Turkish slaves in Malta which were given their liberty on baptism. Malta and Turkey were governed by the same rulers during the Byzantine period, and the slight population affinity seen in the PCA might also be a reflection of this period of political relationship. Another possibility might be that the Phoenician influence on both the Maltese and Turkish populations has acted as an intermediary bringing them together.

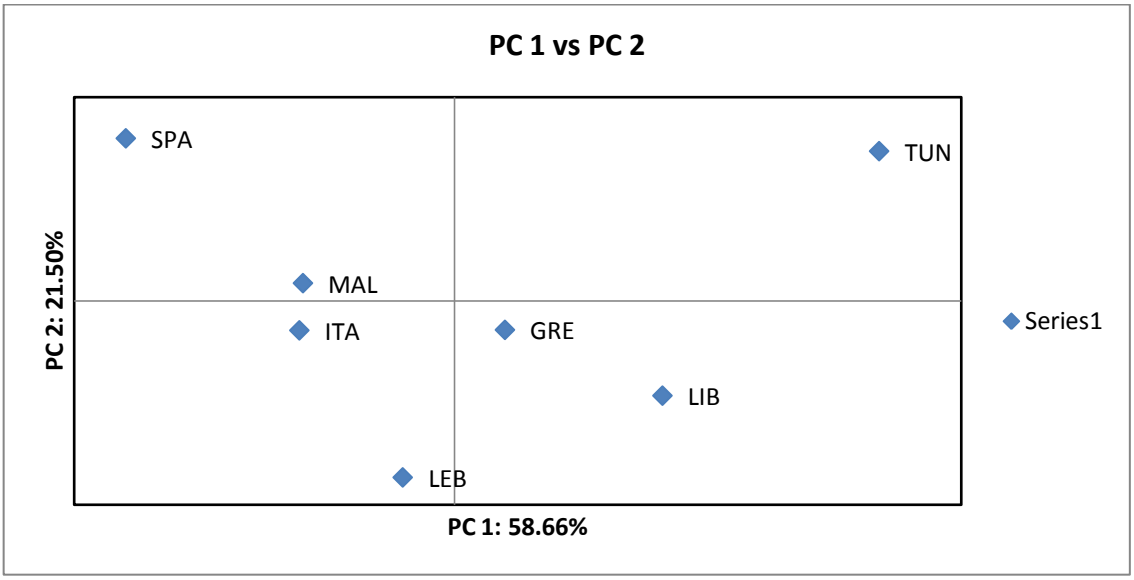

Figure 44: Showing PC 1 vs PC 2 when 8 Y-STR were analysed.

In the analysis involving 8 Y-STRs the picture changes, when PC 1 and PC 2 are compared (Figure 44), which make up for 80.1% of the variation, the Maltese population has Italy as its closest neighbour, with Spain and Greece then following. Of interest in this analysis is that the European Mediterranean populations congregate towards the center except for Spain, with Tunisia, Libya and Lebanon being found on the outskirts of this congregation.

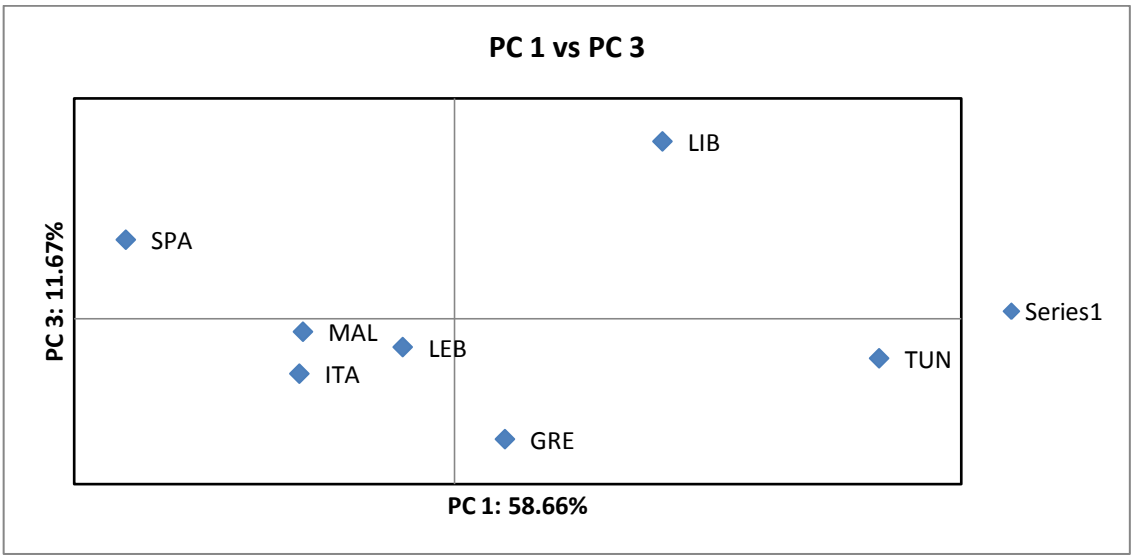

Figure 45: PC 1 vs PC 3 of Mediterranean populations using 8 Y-STRs.

PC 1 vs PC 3 (Figure 45) sees Malta keep its position with Italy being its closest neighbour, but now Lebanon replaces Greece as the closest country to Malta following Italy. PC 2 vs PC 3 (Figure 46) shows once again Malta positioning itself in the middle of the PCA with Italy and Greece being its closest nighbours, whilst the other populations are almost equidistant from the Maltese sample.

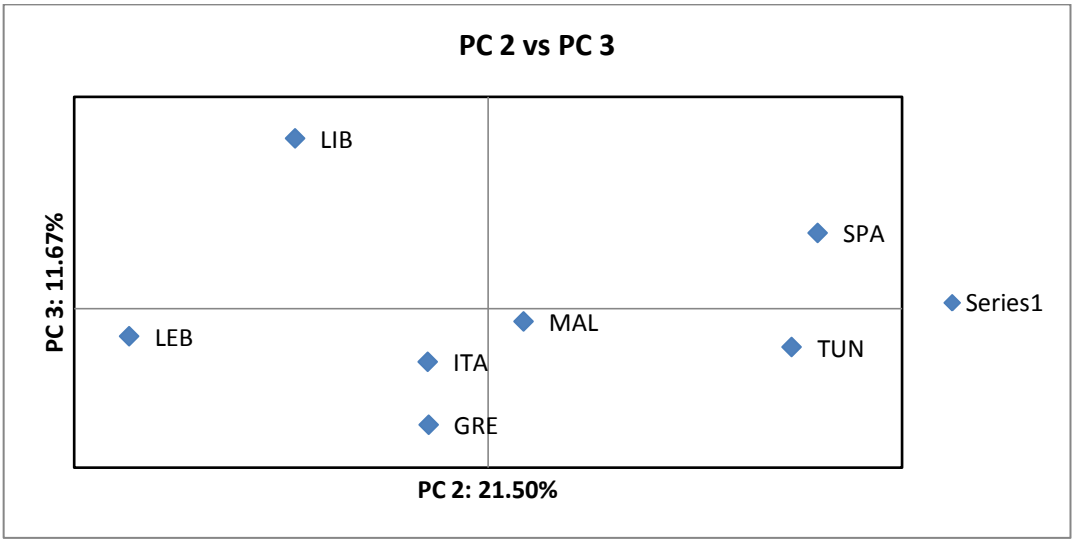

Figure 46: PC 2 vs PC 3 using the 8 Y-STR dataset.

5.5: Y-STR median joining network analysis

Two Median joining networks were constructed in order to observe the relationships between samples from different countries which had identical haplotypes for 5 Y-STRs. This analysis was divided between an Eastern Mediterranean network (Figure 47) and a Western Mediterranean network (Figure 48). The 5 Y-STR loci which were used are DYS19, DYS390, DYS391, DYS392 and DYS393 were chosen as they were the most common loci which were represented in population genetics studies from around the Mediterranean.

Another analysis was conducted using 8 Y-STR loci (DYS19, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438 and DYS439) with a reduced number of populations. Unfortunately the other Y-STRs tested in this analysis had either too much missing data even to try and extrapolate the STRs from the missing data, or the STRs examined in the analysis were not STRs which were found in the papers used as comparison populations in this study. A Y-STR testing kit would have removed these margins of errors, but due to cost considerations the use of a kit was not possible.

A word of caution is that Network (Fluxus technologies) deals with Y-STR data by splitting locuses and thus the networks produced can be very complex, with the same Y-STR showing up as more than one change in the same link. In order to make the networks more visually comprehensable only the number of mutations is shown between one cluster and another. Other ways used to simplify the network visually were that in the analysis only the Maltese sample is fully represented, whilst the other regions were calculated with the Frequency>1 option on. This is the reason the Maltese samples cluster as single haplotypes much more frequently than other populations. The networks were also split into geographic regions with only one of the shortest trees shown in the analysis as the full network makes the analysis incomprehensable due to most clusters linking multiple times to multiple clusters. The full networks can be found in Appendix 11.

The first striking feature of the 5 Y-STR Eastern network (Figure 47) is the number of times that the Maltese samples are found in the same cluster with members of the Lebanese population. This happens 28 times. Whilst conclusions might start being made about them being signs of Phoenician influence, a look at the 8-STR table reveals a different story.

Another interesting feature is that the Maltese population clusters exclusively with the Turkish population several times. This, as mentioned previously, might be a genetic trace of slavery or the influence of the Byzantine empire. Unfortunately the Turkish sample did not have the required STRs to be included in the Eastern 8 Y-STR network (Figure 49), in order to see if this trend continues at a higher resolution.

The Maltese population is found in a cluster with the Sicilian population 17 times in the 5 Y-STR median joining network. This is the same amount of times that the Maltese population clusters with Italian samples in the same analysis. The Sicilian relationship appears to be stronger than the Italian one due to the Sicilian sample being made up from 253 samples when compared to the Italian population which is composed of 292 samples.

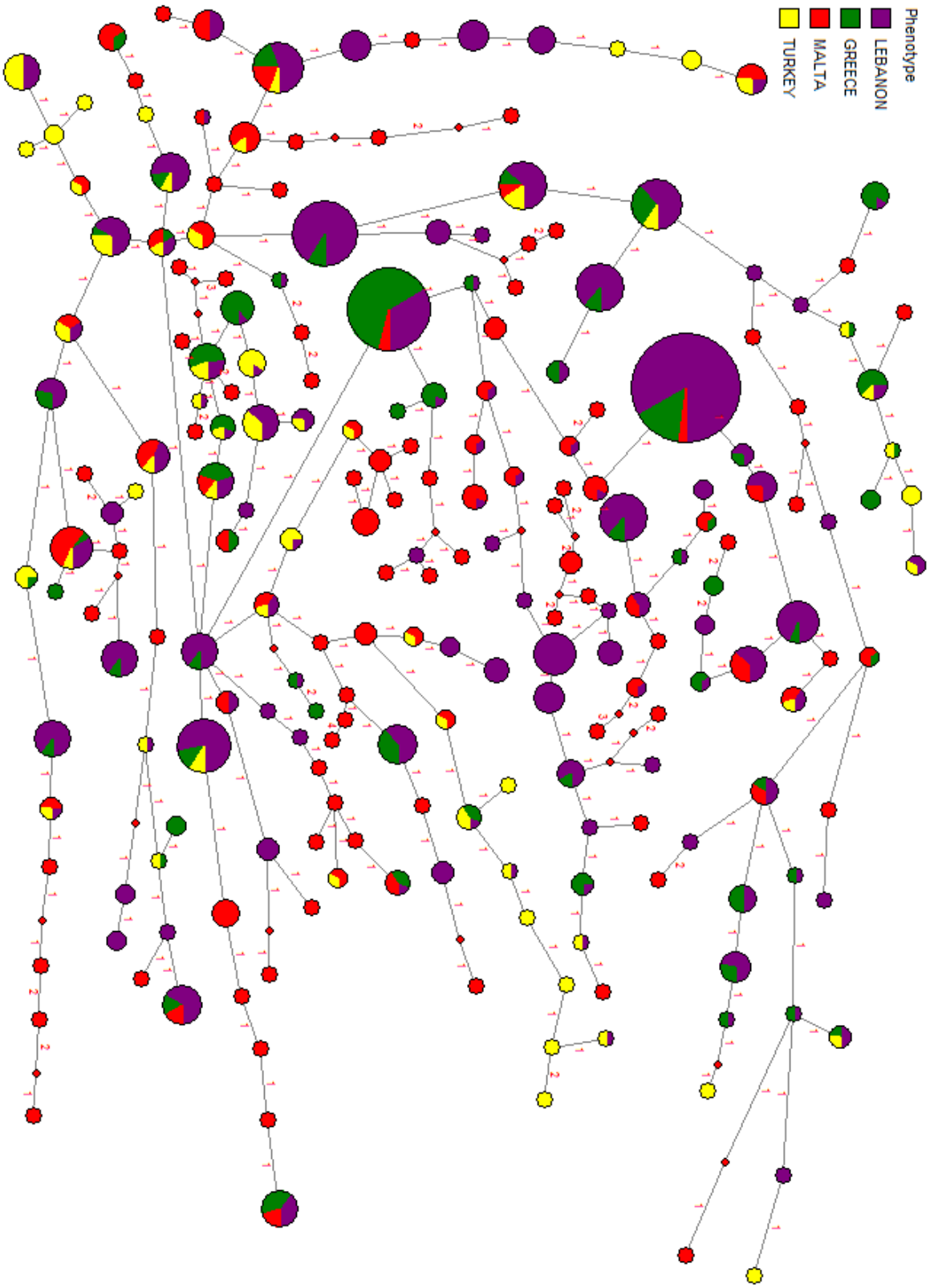

Figure 47: Analysis of Eastern Mediterranean populations using 5 Y-STR.

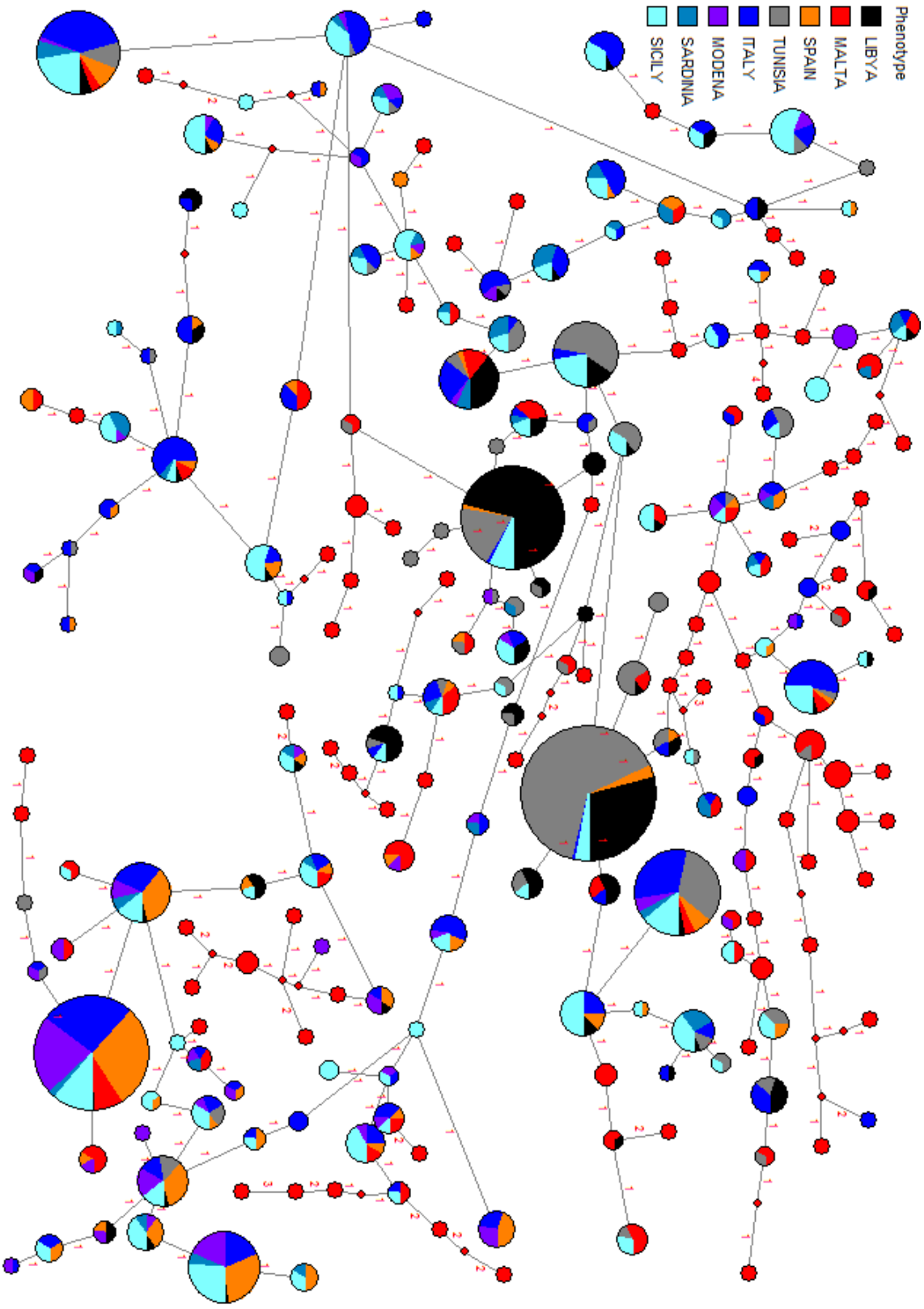

Figure 48: Analysis of Western Mediterranean populations using 5 Y-STR.

In the 8 Y-STR analysis (Figure 49) the Maltese and the Lebanese samples only cluster together six times from the twenty eight times where this happens in the 5 Y-STR analysis.

This indicates that the relationship which might have been suggested by the 5 Y-STR analysis does not in reality exist. Especially when one considers that the Lebanese dataset is the largest dataset in the study with 587 samples. Greek samples cluster with Maltese samples four times in the 8 STR analysis from the thirteen times this happens in the 5 Y-STR analysis, which whilst this might seem less of a link then the Lebanese one, when one remembers that the Greek sample has got 388 samples less in it then the Lebanese one, the link between the Lebanese and the Maltese samples becomes even more tenuous.

The Maltese samples form a cluster with Italian samples nine times in the Western 8 Y-STR network (Figure 50). This is significant when one considers that there are 292 Italian samples in the Network when compared to 587 Lebanese ones. Thus the Italian sample congregates with the Maltese one 1.5 times more than the Lebanese sample, even though it is just under half of the Lebanese populations' sample size.

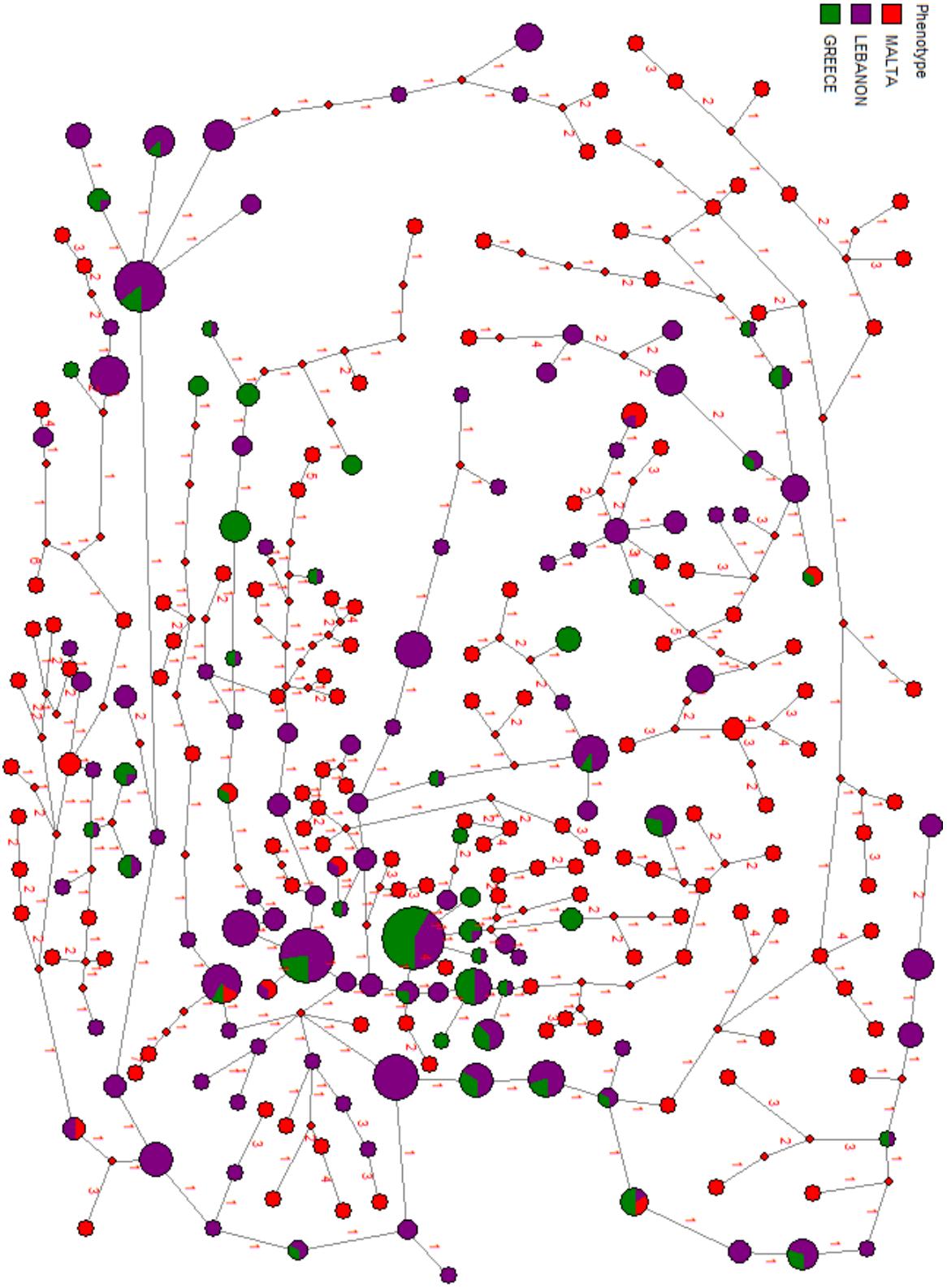

Figure 49: 8 Y-STR Median Joining Network of Eastern populations when compared to the Maltese population.

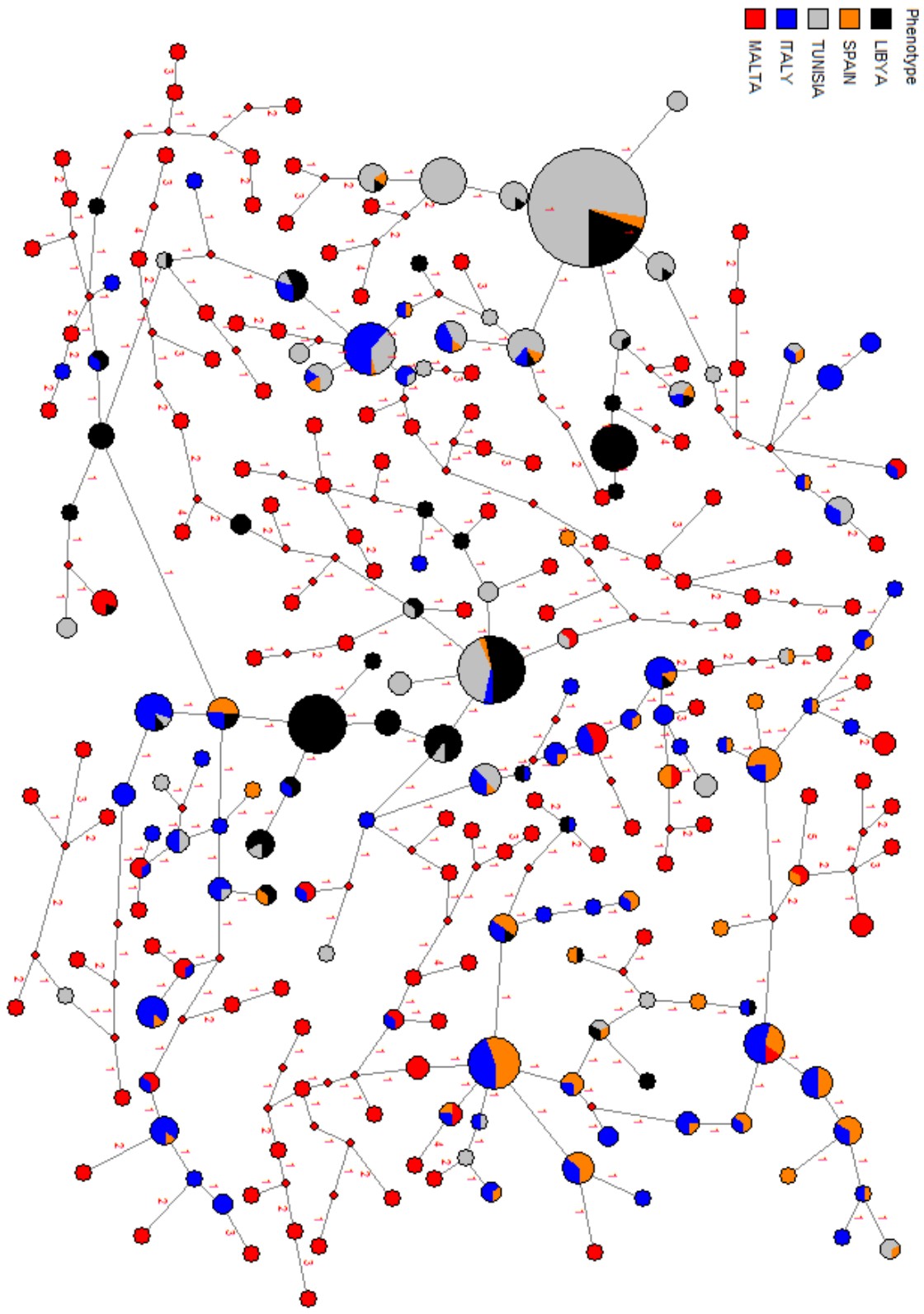

Figure 50: 8 Y-STR Median Joining Network of Western populations when compared to the Maltese population.

The Spanish samples congregate with the Maltese ones four times in the 8 Y-STR analysis when compared to the 15 times in the 5 Y-STR network, whilst the North African samples show a low affinity with the Maltese population, with the Tunisian population congregating 9 times in the 5 Y-STR analysis and the Libyan one 7 times, whilst both populations only cluster one time in the 8 Y-STR analysis. This result is further strengthened when the sample sizes of the populations are considered, with the Tunisian sample being made of 244 individuals and the Libyan sample being made from 239 individuals, which is only a little less than the Italian sample size.

5.6: Comparison of the male and female lineages in the Maltese populations

The results from the 5 Y-STR male lineage study agree with the female lineage results, with the PCA showing Sicily as being the closest population to the Maltese one, and the Aegean also showing influence on the Maltese population. The Italian mainland also shows affinity with the Maltese population both matrilineally and patrilineally.

Of interest is that Lebanon is the country which clusters most often with Malta in the 5 Y-STR patrilineal analysis. Whilst this result might indicate that a Phoenician influence still remains in Malta, which is a theory also supported in the paper by Zalloua *et. al.* (2008) due to the analysis conducted by the researchers on detecting Phoenician expansions in the Mediterranean using seven Y-STRs and linking the findings to a specific haplogroup (J2). When one looks at the 8 Y-STR analysis of this study (Figure 49) this relationship almost disappears completely. The reason for this is that the larger the number of STRs used, the higher the resolution of the study is, and thus relationships which might appear at low resolution might disappear at a higher resolution.

In their study Zalloua *et al.* (2008) mention that the observed relationship between ancient Phoenician sites and Lebanon in terms of Y chromosomal haplogroup J2 can also be seen between these same sites and Greece, and admits that a higher analytical resolution is needed. This pattern, but in terms of STRs is also seen in this analysis, were the observed relationship between Maltese and Lebanese samples dissipates in the 8 Y-STR network whilst apparently being very stong in the 5 Y-STR analysis. Thus the interview given by Wells of the National Genographic project where a claim that more than half the Maltese Y-chromosome lineages could have come in with the Phoenicians (http://ngm.nationalgeographic.com/ngm/0410/feature2/online_extra.html) seems superflous due to the lower genetic resolution the study used, the incorrect attribution of J2 meaning Phoenician when it could have also meant Greek or any expansion from the Near East, and the clustering of Y-STR alleles diffentating by one step together in order to make up for the one-step difference in an STR that it is assumed a 3000 year old haplotype would have accumulated (Zalloua *et al.* 2008).

Whilst the clustering might seem a reasonable solution to the mutation accumulation that samples might experience from the Phoenician period to the modern day, it also means that one will be clustering STRs which might not have been part of the Phoenician migration and expansion. If the same methodology would have been used in the current study, the relationship between the Lebanese population and the Maltese one would have appeared stronger, even though this might not have necessarily reflected reality. The patterns, even if very faint, which one may see in Zalloua *et al.* (2008)'s analysis might not be attributable to the Phoenicians, but might be an expansion that occurred at a different period, either during the Byzantine period or even during the Neolithic or Paleolithic.

What comes as a surprise is that both Libya and Tunisia in both the matrilineal and patrilineal analysis are not as close to the Maltese population as expected. This indicates that the Muslim period in Malta did not have a big genetic impact on the Maltese population.

Chapter 6: Discussion

6.1: The mitochondrial DNA profile of the Maltese population

The mtDNA study was the first such study conducted on the Maltese population. This study showed mtDNA HVR1 affinity between Malta, Sicily and North Italy. The PCA also showed the close affinity of the Maltese population with the Sicilian and South Italian populations. With Sicily and South Italy being always in the top four populations closest to Malta in the PCA analysis of haplogroup frequencies, whilst in the Diversity Information Index PCA the Palermo population sample is always the closest to the Maltese population and the South Italian population again always features in the top four populations closest to Malta in all PCA. The Palermo sample can be considered an extension of the Sicilian sample, with it being representative of a village near Palermo (Alia).

Whilst other populations such as the Egyptian, Libyan, and Palestinian populations are found closer to the Sicilian and South Italian populations in some PC plots, they do not show up consistently in all the PCA as the former two populations do.

The reduced median network haplogroup analysis also shows a close affinity between the Maltese population and the Italian one; with Sicily, North Italy and South Italy being the regions which cluster the most often with the Maltese population. This trend is not followed in haplogroups which are not common in Malta such as X, where then Greece clusters with Malta the most often. The only exception to this rule is the African haplogroup L2 where the Maltese samples cluster with a Cypriot sample instead of samples from North African

populations, as might have been expected due to the predominance of the L haplogroups in the latter populations, and their closer geographic distance to Malta.

A sign of the relatedness between the Maltese and Sicilian populations is the three times that Maltese samples cluster exclusively with Sicilian samples in haplotypes belonging to haplogroup H, which is the most common haplogroup in Europe. This is the only time in the haplogroup analyses where the Maltese samples cluster with one other population only and not with multiple populations.

The closeness of Sicily with Malta confirms the historical, linguistic and archaeological records which have shown contact between Sicily and the Maltese islands from prehistory till the modern day. In their paper, Capelli *et al.* (2005) say that the mtDNA patterns seem to follow the Y chromosome haplotypes in terms of affinity with other populations. This study confirms this prediction, with the Sicilian population being close to the Maltese population as in their Y chromosomal study, although the Sardinian population which clusters close to the Maltese sample in their study does not follow the same pattern in the mtDNA analysis. In the network analysis the Sicilian samples grouped exclusively with Maltese samples several times, thus indicating that the two are related.

The pattern of affinity of the Maltese population with the Greek population although not as strong as the Italian one is still visible, and is probably due to the colonisation of Sicily and South Italy by Greek city states and direct migration from Greek territories to Malta during different periods such as when Malta was under the Byzantine empire. In the reduced median networks the Maltese, Sicilian and Greek samples are found in the same cluster eleven times, although one must note that this never happens exclusively. Previous studies

have also shown the affinity between Sicily and Greece, and this is also one of the reasons why the presence of haplotypes in Malta which are also found in Greece might have come through Sicily. The haplotypes found in the Greek population are also found in Northern and Southern Italy, which is also a possible mode of diffusion for these haplotypes.

A problem with this hypothesis is that there are a number of times were the Greek samples cluster with the Maltese ones without the Sicilian or mainland Italian samples being present in the cluster. Sometimes this clustering of Maltese and Greek samples also happens exclusively. A possible period for the arrival of the Greek influence on the island of Malta is the Byzantine period, although even here the main cultural influence on the island was still its closest neighbour Sicily (Dalli 2006). At the same time any attempt to date such migrations without the use of a molecular clock is superfluous.

In the PCA, the Tunisian, Libyan and Moroccan samples were rarely as close to the Maltese samples as expected when their geographic position in relation to Malta is considered. This might indicate that the Arab period in Malta did not leave such a strong genetic impact on the Maltese islands as might be thought, especially when one considers that the significant presence of L2 and L3 haplogroups which are present in the Maltese population do not cluster with any North African samples.

The presence of these African haplogroups in Malta also puts in doubt Pereira *et al.* (2000)'s argument that the L haplogroups in Portugal are due to the Sub-Saharan slave trade whilst the U6 haplogroup present is due to the Moorish invasions. In Malta the L haplogroups are present at significant levels even though historically most slaves in Malta were of Near Eastern origin and not from Sub-Saharan Africa.

It is known that in the case of Malta foreign women, especially Sicilian ones, did settle in Malta and intermingled with Maltese males during the period of the Knights of Saint John (Vella 1979). An interesting episode of the Maltese islands was that in 1565 just before the Great Siege, a large number of the women, as well as the weak and the old of the islands, were transferred to Sicily so that they would not be a burden for the fighting men on the island. Some of the families which left during that time did not return to the island after the siege. These two episodes might be part of the explanation of the presence of shared haplotypes between the two islands. Even though these shared haplotypes could have occurred at any point during history or prehistory.

The Lebanese sample does not seem to cluster with the Maltese samples more often than other populations do, especially when its population size of 363 samples is considered. This sample size is almost the same as that of North Italy (395), but the North Italian samples cluster with Maltese samples more often than the Lebanese ones do. The Lebanese population always clusters with Malta when other populations are also present, and never exclusively.

6.2: Y-STR analysis of the Maltese population

In order to save costs, a modification of the standard genotyping method was used, in which the PCR products were labelled via a generic forward primer, to which the dye was attached, rather than using individual labelled primers for each STR. Whilst the method gave good results, it would not be suitable for forensic DNA labs as there was a high failure rate, which is the reason why the results for some STRs are not analysed here. This could have been due

to the need for more optimisation of the primers. The problem with the method is that pre-screening cannot be done with agarose gel electrophoresis and thus only the genotyping itself shows whether the procedure has worked. This increased the cost of optimisation considerably and thus some STRs which were originally included in the study were not examined further after initial failures. Another issue with the method is that multiplexing such a system is difficult, and the attempts to do so in this study failed due to the generic primer attaching preferentially to some of the tailed specific primers. It was not deemed cost efficient to try and optimise such a system for this study even though attempts were made. As a proof of concept for a cost-effective method of human STR analysis the system still gave adequate results, which were corroborated by replication and by comparing the results obtained in this study with other results involving the Maltese population using traditional genotyping methods.

The higher rate of observed haplotype variability in the male Maltese population indicates that the Maltese male genetic line is more variable than the female genetic line, and thus that more males migrated in and out of the population, bringing with them this variability. This phenomenon is also observed in the historical records where Mercieqa (2006) studied the number of marriages that occurred between Maltese and Venetians over the period of the Knights of Saint John. The results showed that there were instances where a large number of mixed marriages were occurring. These marriages always involved Venetian males marrying Maltese females. Even a low rate of these kind of marriages, when observed over a series of decades, will affect the Y chromosome gene pool of the population. The historical mixed marriages between British males and Maltese females will also have affected the variability. During World War 1, when the Japanese navy was berthed in Malta, records also exist of Maltese females marrying men from as far away as Japan (Ganado 1977).

When a PCA was conducted with the 5 Y-STRs using all the populations under study, the results agreed with the findings of Capelli *et al.* (2005), which showed that the Maltese population has the closest genetic affinity with the Sicilian populations. After this population the Italian one is also the closest to the Maltese population. These two populations change interchangeably amongst the different PC in which one of them comes as the closest to the Maltese one. The interesting thing in the present study is that in the 8 Y-STR PCA whilst the Italian population always is the closest one to the Maltese population, the Greek and the Lebanese populations are the closest after. Whilst the relationship with the Greek population can be explained due to migrations from Sicily and Greece itself, the affinity with the Lebanese population is more problematic to explain. The most logical explanation would be that there are still artefacts from the time of the Phoenicians. This is interesting as Zalloua *et al.* (2008) also mention Malta as one of the regions where the Phoenicians left their trace in the modern world. Whilst this study could not identify specific 'Phoenician' haplotypes, the link between the two populations was visible in the 5 Y-STR network analysis. This apparent link then disappears in the 8 Y-STR analysis, thus showing that at a higher resolution no clear relationship is visible between the Maltese and the Lebanese populations. The link observed by Zalloua *et al.* (2008) is tenuous due to the way the samples were analysed as was previously mentioned in the Y chromosome part of the results chapter. The fact that Lebanon comes close to European populations in the PCA might be an indication of the influence the Byzantines and the crusaders had on the population as is also mentioned by Zalloua *et al.* (2008).

Both in the 5 and in the 8 Y-STR PCA the Tunisian and Libyan samples are never positioned close enough to the Maltese ones to indicate that they have contributed

extensively to the Maltese population. This is unexpected considering the close geographic proximity and the period of Arab domination in Malta, during which no signs of Christianity can be found on the island, thus indicating that the whole population was Muslim (Luttrell 2002).

Whilst studies concerning the Maltese surnames indicated that after the Norman invasion in Malta the local Maltese population started giving themselves surnames and these were invariably Semitic, whilst the people who settled in with the Normans had more Latin surnames, this does not mean that the Maltese on the island at that time were of North African descent. Another episode in history which shows that the Maltese might not have intermingled with their Arab occupiers occurred when there was an attempt to retake Malta by the Byzantine forces, and the local rulers had to promise the Maltese population that they would be treated equally in order to gain their support (Dalli 2006). This is of interest as it shows that the dominating force did not integrate fully with the local population.

The fact that in the early Middle Ages certain church rites are known by their Greek names instead of the Latin ones, even though Malta was a Latin rite church, indicates that these words had been used on the island before the Arab invasion and had then been adapted during that period and then returned to their original purpose with the advent of the Normans (Buhagiar 2002). This word adaptation can also be seen in modern Maltese where the name for Lent in Maltese is ‘Randan’, which although 98% of the Maltese population is Catholic still hails from the time of the Muslim period, when the word for the Muslim version of Lent is ‘Ramadan’.

An example of this lack of integration from history books is found when the Maltese petitioned the Sicilian curia during the Medieval period to remove a priest who had been sent from Sicily to Malta to teach. The reason the Maltese governing body gave for this was that the priest could not speak Maltese and thus it was difficult for the locals to interact with him. This language barrier might have proved a stumbling block for the lower class Maltese to integrate with foreigners, as the only language they could speak was Maltese and only the nobility was able to speak other languages. Another example of the lack of integration concerns the last Grandmaster of the Knights of Saint John in Malta, Ferdinand Von Hompesch, who was well loved by the Maltese as he was the only Grand Master who was able to speak the native language (Blouet 1976). Thus during the time of the Knights of Saint John it had taken over 250 years until one of their leaders could effectively communicate with the population. This highlights the barrier that there was between the Maltese general population and the ruling class. The Maltese were much more easily integrating with Sicily through commerce, which can be seen linguistically from the Sicilian words which were entering the Maltese language at this time. This integration with Sicily is also seen genetically with the closeness of the two populations both from the matrilineal and the patrilineal perspectives.

The fact that Malta is not close to Libya or Tunisia genetically indicates that throughout Maltese history there was no period when Malta was depopulated. This issue has been debated extensively due to the problems arising between the Byzantine and the Arab period, when historical records are few and it was not known whether the islands had been abandoned. The genetic evidence indicates that the Maltese islands were never depopulated, since if the islands would have been repopulated during the Islamic expansion more North African links would have been found with the Maltese population.

6.3: Ancient DNA analysis and its suitability in Malta

The aDNA analysis was partly successful in obtaining sequences from Maltese archaeological remains. This was the first time that Maltese archaeological remains had been submitted to such an analysis. From this study it can be concluded that there is strong evidence that DNA can survive in the Maltese climate for centuries, although the chances of amplifying the DNA become less as time goes by. This is exemplified from the difference between the Tal-Gardina and the Tal-Barrani samples where these sites produced more sequences and had a higher replication rate than the samples from Kercem. There is doubt whether the sequences amplified from the Kercem samples were true aDNA. Whilst the other sites suffered from a high rate of contamination due to the excavation methodology and the conditions they were kept in until the analysis, making even their results slightly doubtful . What indicates that aDNA was truly amplified are the instances were DNA was obtained from samples which showed mutations not present in the modern Maltese population, such as in the case of TG115B whose amplification were consistent with haplogroup U6 which is not present in the modern population. The mutation motif was formed by overlapping sequences obtained from different amplifications, some of which were also successfully replicated from a separate sample of the same inhumation.

A maternal relationship between burials could not be established at the Tal-Barrani site as all the polymorphisms in the commonly amplified regions were different. Of interest is sample BR1 which is probably a female. Strength is added to the probability that this result is correct as there is less chance of allelic dropout having occurred if the sample was contaminated by the male analyser, as this would have given both X and Y amplicons since

the contaminant would have been modern male DNA. This indicates that women were buried in the Tal-Barrani site, even though the bone physiology had indicated that the persons buried there were used to hard labour, which is normally associated with males. On the other hand the possibility of a female contaminant cannot be excluded.

In the Tal-Gardina study a number of sequences that were obtained corresponded with the CRS. This is problematic as some of the archaeologists who excavated the site have the CRS in the amplified part of the aDNA. Thus these sequences had to be discarded as possible contaminants even though this is the most common haplogroup in Europe and might have been an indication that the inhumations were related.

The Kercem site unfortunately only gave two successful amplifications from a total of six samples from three inhumations, and the sequences that were obtained could not be replicated. One of the successful amplifications was in HVR2, the polymorphism that was present showing that it is not haplogroup H, thus making it different from the most common haplogroup found in Europe. The other sequence could not be assigned a haplogroup as the combination of polymorphisms is not known in modern populations, and parts of the sequence looks like the product of artefacts.

The limited success with the Kercem samples indicates that sites of this age might be at the threshold for aDNA analysis on the Maltese islands, using the classical PCR amplification methods if what was amplified was truly ancient DNA, of which there is doubt.

Of note was the fact that the inhumations found at Kercem had been flooded with water and it took over a year to dry the site before it was excavated. Counteracting this reason for the

limited DNA preservation is the fact that the aDNA analysis was conducted on the Kercem samples only a few months after their excavation, instead of the decades for the Tal-Barrani samples and the few years with the Tal-Gardina samples. This means that for the samples from Kercem post-excavation degradation was not as substantial an issue is it might have been for the Tal-Barrani and the Tal-Gardina samples.

6.4: Conclusions derived from this study

Whilst the method of STR genotyping used in this study served its purpose in halving the cost it would have taken for traditional genotyping methods to be conducted, the method was not very accurate, with several samples having to be retyped in order to produce a result and some loci failing even after the second attempt. Another factor is that the method will be slow to use unless a multiplex is developed. This study's attempts at developing such a multiplex failed due to the non-specificity of the labelling method, thus making a three- or four-dye genotyping multiplex not possible. The conclusion is that the method is useful for genotyping STRs cheaply but the payoff is that more time is spent in the lab and analysis and interpretation of the data is more difficult.

The Maltese population has been influenced by a mix of different populations which we know they have been in contact with from Maltese history, and this is reflected in the archaeological, linguistic and now also genetic records. Whilst the greatest influence remains Italy in the formulation of the Maltese population, other influences such as Greece indicate that Malta did not look solely to the Italian mainland for colonisations. The presence of mitochondrial haplogroup L2a also indicates an African character to the Maltese

population, even though the haplotypes found in Malta are were not present in the modern day Tunisian, Moroccan and Libyan samples, but cluster with a Cypriot sample.

The aDNA analysis has shown that such studies are possible in Malta even though the information garnered from them might be limited until new technologies currently being developed to study aDNA are applied to them. Whilst it may not be as effective as in countries with a colder climate, aDNA studies in the Maltese islands might still be useful to draw certain conclusions about the Maltese population of the period and also about kinship aspects on particular sites if the sites are excavated and investigated in a way to lower the chance of contamination, and the samples are kept at low temperatures to reduce post-excavation DNA degradation.

Autosomal and Y chromosomal aDNA amplifications will not be as successful as mtDNA studies with Maltese archaeological samples, and should not be a focus of work on the Maltese archaeological record. Another aspect of aDNA analysis which seems to play an important role in Malta is the question of DNA contamination. Whilst nothing can be done if the excavator of the site and the inhumation share the same haplotype and thus are undistinguishable, it is important to always strive to see whether it was really contamination or not. All the CRS sequences obtained from the ancient Maltese samples had to be discarded in order to make sure that false positive results were not included in the analysis. With the proper clothing and preventative measures whilst excavating, the archaeologists could minimise the worries of false positives. Another mode of practice which can be introduced is that the person who is going to do the aDNA analysis is the only person who excavates an inhumation and thus comes into contact with it before the samples for aDNA analysis is taken. This would minimise the problem of having to test several people in order

to see if their haplotypes match the sequences obtained from the samples. It would also avoid the possibility of having to discard what might be authentic aDNA results if one of the excavators shares the same haplotype coincidentally.

Another route that can be taken in Maltese archaeology for the time being is that of only working on non-human specimens until the technology improves and becomes cost efficient enough to try aDNA analysis with the newer emerging technologies. The reason why the search for non-human DNA is seen as more feasible than human DNA is that any contamination by the researchers or the excavators is immediately identified as human and thus can be discarded. Another factor that would help aDNA studies in Malta would be an effort to keep any recovered material which has been identified as suitable for aDNA analysis in a -20^{o}C environment until the time when the analysis is conducted. This would reduce post-excavation DNA degradation considerably and thus increase the quality of the analysis.

6.5: Future work

A study comparing the Y chromosome STR data with specific surnames of the Maltese population would be useful in order to try and determine whether certain surnames can be identified as coming from certain countries genetically. The reason for this is that Maltese surnames broadly fall under the Semitic or the Latin form, and it would be interesting to see if this divide in surnames can be also expressed genetically. A study with a greater number of alleles would also improve the resolution of the results obtained in this study.

A mitochondrial analysis which encompasses SNPs from all over the mitochondrial genome would be more precise in determining the haplogroup of samples than just using HVR1 and HVR2 alone. Complete mitochondrial sequencing would also be useful in studying the Maltese population as such a study would increase the resolution and be able to discern genealogies better in the Maltese population. Complete mitochondrial DNA sequencing also has the benefit of being able to evaluate the precision of analysis based on the study of HVR1 and 2 or RFLP data (Richards and Macaulay 2001).

High-resolution SNPs for the Y chromosome would also be beneficial in identifying patterns between the Maltese population and other populations from around the Mediterranean which the present study has not discerned.

The best way to look at different populations would be by comparing whole genomes of samples from different populations. This is the best way as it involves the highest resolution possible, but its huge limitation is the cost involved, although with complete genome sequencing prices lowering all the time, it might be possible to conduct such a study in the near future without needing an exorbitant budget.

Ancient DNA analysis should be conducted with samples from several periods of the Maltese archaeological record whilst noting the burial conditions and the time between excavation and analysis, starting from the modern period and progressively going back until aDNA amplification is not possible. This would help improve the benchmark set in this study for aDNA analysis, and the observation of the burial conditions would help future researchers in identifying which samples are the most suitable for aDNA analysis.

As the number of aDNA studies from around the Mediterranean increases, the Maltese ancient samples should be compared to these studies, in order to try and determine migration patterns in the ancient world. The Maltese samples should also be subjected to stable isotope analysis in order to determine whether these persons had lived in Malta all their life or whether they had migrated from elsewhere.

Ancient DNA analysis on plant and animal material from the Maltese archaeological record should also be conducted as this would be useful in identifying which flora and fauna were prevalent in Malta throughout different periods and how these changed with time. This is needed due to the deforestation that Malta is known to have suffered since the Neolithic and the fact that the difference in goat and sheep bones which are also found in the Maltese archaeological record cannot be determined through traditional osteological studies.

Another interesting study would be to try and explore if a genetic relationship exists between Malta and Rhodes. The reason for this is that both islands were ruled for centuries by the Knights of Saint John, and thus any male descendants they might have left could possibly be identified as such. The reason is that the Knights came from a small group of noble families, which is why when one looks at the Knights throughout their history, the same surnames continue cropping up. It would be interesting to trace the male descendents of these noble families and compare them to the males from Malta and Rhodes in order to see if any relationship may be determined.

A final study which would yield interesting results would be the genetic analysis of old people from different villages in Malta. Whilst nowadays people in Malta go and live in other villages regularly, until a few decades ago this was a very rare occurrence. This gives

the opportunity to try and determine whether different villages had different genetic patterns in the Maltese islands, with some showing more genetic diversity than others. Unfortunately, this study must be conducted shortly if meaningful results are to be obtained from it, as with the older generation dying out and the younger one spreading out all over Malta, any discernible influences might soon be lost. One can also do this by asking the people being tested the village of origin of their grandparents but this might not be as accurate due to the migrations that occurred internally in Malta during WW2 where people moved from the areas around the Grand Harbour; which suffered the worst of the bombing, to safer places in Malta were a lot of them then settled. This might mean that the descendents might not know about this migration dating to seventy years before.

References:

Abela, F., 1647. *Della descrizione di Malta*. Stamperia di S.A.S. Malta.

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., Loogvali, E.-L., Kivisild, T., Bandelt, H.-J., Richards, M., Villems, R., Santachiara-Benerecetti, S., A., Semino, O., Torroni, A., 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *American Journal of Human Genetics* **75**: 910-918.

Addinsoft, 2007. *XLSTAT, Analyse de donnees et statistique avec MS Excel*. Addinsoft. USA.

Agilent, 2008. StrataClone *PCR cloning kit instruction manual*. Agilent Technologies Inc. USA.

Anderson, S., Bankier, A. T., Barrell, B. G., De Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., Young, I. G., 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290**: 457–465.

Appendix, 1997. Reference tables to: Evaluation of Y-chromosomal STRs: a multicenter study (Kayser, *et al*.) and chromosome Y microsatellites: population genetic and

evolutionary aspects (de Kniff, *et al*.). *International Journal of Legal Medicine* **110**:141-149.

Applied Biosystems, 2010**.** *GeneScan® Reference Guide, Chemistry Reference for the ABI PRISM®310 Genetic Analyzer*. Applied Biosystems. USA.

Ayub, Q., Mohyuddin, A., Qamar, R., Mazhar, K., Zerjal, T., Mehdi, S. Q., Tyler-Smith, C., 2000. Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Research* **28**: e8.

Bandelt, H-J., Forster, P., Sykes, B. C., Richards, M. B., 1995. Mitochondrial portraits of human populations using Median Networks. *Genetics* **141**: 743-753.

Bandelt, H-J., Forster, P., Röhl A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology Evolution* **16**:37-48.

Biotium, 2009. *GelRedTM & GelGreenTM Enviornmentally safe and ultra-sensitive nucleic acid gel stains for replacing EtBr*. Biotium. USA.

Blouet, B., 1976. *The story of Malta*. Faber Paperbacks. London.

Bonanno, A., 1992. *Roman Malta: The archaeological heritage of the Maltese islands*. World confederation of salesian past pupils of Don Bosco. Rome.

Bosch, E., Calafell, F., Perez-Lezaun, A. 1997. Population history of North Africa: evidence from classical genetic markers. *Human Biology* **69**: 295-311.

Bosch, E., Calafell, F., Perez-Lezaun, A. 2000. Genetic structure of north-west Africa revealed by Y-STR analysis. *European Journal of Human Genetics.* **8**: 360-366.

Bosch, E., Calafell, F., Comas, D., Oefner, P. J., Underhill, P. A., Bertranpetit, J., 2001. High-Resolution analysis of human Y-Chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *American Journal of Human Genetics* **68**: 1019-1029.

Bouwman, A.S., Brown, K.A., Prag, A.J.N.W., Brown T.A., 2008. Kinship between burials from Grave Circle B at Mycenae revealed by ancient DNA typing. *Journal of Archaeological Science* **35**: 2580-2584.

Bouwman, A.S., Chilvers, E.R., Brown, K.A., Brown, T.A., 2006. Identification of the authentic ancient DNA sequence in a human bone contaminated with modern DNA. *American Journal of Physical Anthropology* **131**: 428-431.

Brincat, J., 1991. *Malta 870-1054: Al-Himyari's account and its linguistic implications*. Said International. Malta.

Brody, J. R., Kern, S. E., 2004. Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. *BioTechniques* **36**: 214-216.

Brown, T. A. 2002. *Genomes*. Second edition. BIOS Scientific publishers Ltd. Oxford.

Brown, T. A., Brown, K. A., 1992. Ancient DNA and the archaeologist. *Antiquity* **66**: 10-23.

Brown, T. A., Brown, K. A., 2011. *Biomolecular Archaeology: An introduction*. John Wiley & Sons. USA.

Buhagiar, M., 2002. The re-christianization of Malta: Siculo-Greek monasticism, Dejr toponyms, and rock-cut churches. *Melita Historica* **13**: 253-284.

Burger, J., Hummel, S., Herrmann, B., Henke, W., 1999. DNA preservation: A microsatellite-DNA study on ancient skeletal remains. *Electophoresis* **20**: 1722-1728.

Butler, J. M., Schoske, R., Vallone, P. M., 2002. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Science International* **129**: 10-24.

Butler, J. M., Schoske R., 2004. Forensic value of the multicopy Y-STR marker DYS464. *International Congress Series* **1261**: 278-280.

Butler, J. M., 2005. *Forensic DNA Typing*. Second edition. Elsevier Academic Press. London.

Calo, C. M., Melis, A., Vona, G., Piras, I.S., 2008. Sardinian Population (Italy): a genetic review. *International Journal of Modern Anthropology* **1**: 1-121

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

Capelli, C., Redhead, N., Romano, V., Cali, F., Lefranc, G., Delague, V., Megarbane, A., Felice, A. E., Pascali, V. L., Neophytou, P. I., 2005. Population structure in the Mediterranean Basin: A Y Chromosome Perspective. *Annals of Human Genetics* **70**: 207-225.

Carlos Flores, C., Maca-Meyer, N., Gonzalez, N. M., Oefner, P. J., Shen P., Perez, J. A., Rojas A., Larrug, J. M., Underhill, P.A., 2004. Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *European Journal of Human Genetics* **12**: 855-863.

Cassar, M., Farrugia, C., Vidal C., 2008. Allele frequencies of 14 STR loci in the population of Malta. *Legal Medicine* **10**: 153-156.

Cavalli-Sforza, L.L., Edwards A. W. F., 1967. Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* **19**: 233-257.

Champlot, S., Berthelot, C., Pruvost, M., Bennett, A., Grange, T., Geigl, E.M., 2010. An Efficient Multistrategy DNA Decontamination Procedure of PCR Reagents for Hypersensitive PCR Applications. *PLOS ONE* 10.1371/journal.pone.0013042

Cherni, L., Pereira, L., Goios, A., Loueslati, B.Y., Khodjet el Khil, H., Gomes, I., Gusmão, L., Alves, C., Slama, A., Amorim, A., Elgaaied, A.B., 2005. Y-chromosomal STR haplotypes in three ethnic groups and one cosmopolitan population from Tunisia. *Forensic Science International* **11**: 95-99.

Chilvers, E. R., Bouwman, A. S., Brown, K.A., Arnott, R.G., Prag, A. J. N. W., Brown, T. A., 2008. Ancient DNA in human bones from Neolithic and Bronze Age sites in Greece and Crete. *Journal of Archaeological Science* **35**: 2707-2714.

Coble, M. D., Butler, J. M., 2005. Characterization of New MiniSTR loci to aid analysis of degraded DNA. *Journal of Forensic Science* **50**: 43-53.

Colantonio, S. E., Lasker, G. W., Kaplan, B. A., 2003. Use of surname models in human population biology: A Review of Recent Developments. *Human Biology* **75**: 785-807.

Cooper, A., Poinar, H. N., 2000. Ancient DNA: Do it right or not at all. *Science* **289**: 1139.

Côrte-Real H. B, Macaulay V. A., Richards M. B, Hariti, G., Issad M. S., Cambon-Thomsen, A., Papiha, S., 1996. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Annals of Human Genetics* **60**: 331-350.

Dalli, C., 2006. *Malta: The Medieval millenium*. Midsea Books Ltd. Malta.

Di Gaetano, C., Cerutti, N., Crobu, F., Robino, C., Inturri, S., Gino, S., Guarrera, S., Underhill, P. A., 2009. Differential Greek and Northern African migrations to Sicily are

supported by genetic evidence from the Y chromosome. *European Journal of Human Genetics* **17**: 91–99.

Di Rienzo, A., Wilson, A.C., 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Evolution* **88**: 1597-1601.

Elmrghni, S., Coulson-Thomas, Y. M., Kaddura, M., Dixon, R. A., Williams, D. R., 2012. Population genetic data for 17 Y-STR markers from Benghazi (East Libya). *Forensic Science International Genetics.* **6**: 224-227.

Epoch Life Science. 2008. GenCatchSpin™ All-in-1 Mini Spin Columns for DNA/RNA extraction. Epoch Life Science. USA.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47-50.

Fabrini, E., 2005. mtDNA. http://www.doppiovu3.it/mtdna/ Last accessed 26/10/2012

Fadhlaoui-Zid, K., Rodriguez-Botigue, L., Naoui, N., Benammar-Elgaaied, A., Calafell, F., Comas, D., 2011. Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile valley. *America Journal of Physical Anthropology* **145**: 107-117.

Felsenstein, J., 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**: 164-166.

Flores, C., Maca-Meyer, N., Gonzalez, A. M., Oefner, P. J., Shen, P., Perez, J. A., Rojas, A., Larruga, J. M., Underhill, P. A., 2004. Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *European Journal of Human Genetics* **12**: 855-863.

Fluxus Engineering., fluxus-engineering.com last accessed 14/10/2011

Forster, P., Röhl, A., Lünnemann, P., Brinkmann, C., Zerjal, T., Tyler-Smith, C., Brinkmann, B., 2000. A Short Tandem Repeat-Based phylogeny for the human Y chromosome. *American Journal of Human Genetics* **67**:182-196.

Fraumene, C., Petretto, E., Angius, A., Pirastu, M., 2003. Striking differentiation of sub-populations within a genetically homogeneous isolate (Ogliastra) in Sardinia as revealed by mtDNA analysis. *Human Genetics* **114**: 1-10.

Ganado, H., 1977. *Rajt Malta tinbidel*. Interprint, Malta.

Gerstenberger, J., Hummel, S., Schultes, T., 1999. Reconstruction of a historical genealogy by means of STR analysis and Y-haplotyping of ancient DNA. *European Journal of Human Genetics* **7**: 469-477.

Ghiani, M.E., Vona, G., 2002. Y-chromosome specific microsatellite variation in a population sample from Sardinia (Italy). *Collegium antropologicum* **26**:387-401.

Ghirotto, S., Mona, S., Benazzo, A., Paparazzo, F., Caramelli, D., Barbujani, G., 2010. Inferring genealogical processes from patterns of Bronze-Age and modern DNA variation in Sardinia. *Molecular Biology and Evolution* **27**: 875-886.

Gonzalez, A. M., Karadsheh, N., Maca-Meyer, N., Flores, C., Cabrera, V. M., Larruga, J. M., 2008. Mitochondrial DNA variation in Jordanians and their genetic relationship to other Middle East populations. *Annals of Human Biology* **35**: 212-231.

Guglielmino, C. R., Zei G., Cavalli-Sforza L. L., 1991. Genetic and cultural transmission in Sicily as revealed by names and surnames. *Human Biology* **63**: 607-627.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium* **41**: 95-98.

Hedges, S. B., Kumar, S., Tamura, K., Stoneking, M., 1992. Human origins and analysis of mitochondrial DNA sequences. *Science* **255**: 737-739.

Höss, M., Jaruga, P., Zastawny, T. H., 1996. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Research* **24** 1304-1307.

Invitrogen, 2006. *TOPO® TA Cloning User Manual*. Invitrogen. USA.

Irwin, J., Saunier, J., Strouss, K., 2008. Mitochondrial control region sequences from northern Greece and Greek Cypriots. *International Journal of Legal Medicine* **122**: 87-89.

Jorde, L. B., Bamshad, M., Rodgers, A. R., 1998. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *BioEssays* **20**: 126-136.

Kayser, M., Kittler, R., Erler, A., 2004. A comprehensive survey of human Y-chromosomal microsatellites. *American Journal of Human Genetics* **74**: 1183-97.

Kidd, D. M., Ritchie, M., G., 2006. Phylogeographic information systems: putting the geography into phylogeography. *Journal of Biogeography* **33**: 1851-1865.

Kirsanow, K., Burger, J., 2012. Ancient human DNA. *Annals of Anatomy* **194**: 121-132

Kloss-Brandstaetter A., Pacher D., Schoenherr S., Weissensteiner H., Binna R., Specht G., Kronenberg F. 2010. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups http://www.haplogrep.uibk.ac.at doi: 10.1002/humu.21382

Kouvatsi, A., Karaiskou, N., Apostolidis, A., Kirmizidis, G., 2001. Mitochondrial DNA sequence variation in Greeks. *Human Biology* **73**: 855-869.

Kovatsi, L., Parsons, T. J., Just, R., S., Irwin, J. A., 2006. Genetic variation for 15 autosomal STR loci (PowerPlex 16) in a population sample from northern Greece. *Forensic Science International* **159**: 61-63.

Kovatsi, L., Saunier, J. L., Irwin, J. A. 2009. Population genetics of Y-chromosome STRs in a population of Northern Greeks. *Forensic Science International Genetics* **4**: 21-22.

Krings, M., Stone, A., Schmitx, R. W., Krainitzki, H., Stoneking, M., Pääbo, S., 1997. Neanderthal DNA sequences and the origin of modern humans *Cell* **90**: 19-30

Librado, P. and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451-1452.

Li, R. C., 2009. *Application of proteinases for DNA isolation of bone specimens*. NCJRS. Research report submitted to the U.S. Department of Justice. Document No: 227502.

Livy., *Hannibal's war books 21-30*. Translation by J.C. Yardley. Edt. Hoyos, D. Oxford University Press, Oxford.

Luttrell, A., 1975. *Medieval Malta: Studies on Malta before the Knights*. The British School at Rome, London.

Luttrell, A., 2002. *The making of Christian Malta: From the Early Middle Ages to 1530*. Ashgate Variorum, Hampshire.

Maca-Meyer N., Sanchez-Velasco, P., Flores C., Larruga J.M., Gonzalez A., Oterino A., Leyva-Cobian, F., 2003. Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). *Annals of Human Genetics* **67**: 329-339.

Mallia-Milanes, V., 1988. *The British colonial experience 1800-1964*. Mireva Publications. Malta.

Malone, C., Bonanno, A., Guder, T., Stoddard, S., Trump, D., 2005. The death cults of prehistoric Malta. *Scientific American* **15**: 14-23.

Malone, C., Stoddart, S., Bonanno, A., Trump, D.(ed.) Gouder, T., Pace, A., 2009 *Mortuary customs in prehistoric Malta: excavations at the Brochtorff Circle at Xagħra (1987–94)* McDonald Institute Monographs. Cambridge.

Marota, I., Basile, C., Ubaldi, M., 2002, DNA decay rate in papyri and human remains from Egyptian archaeological sites. *American Journal of Physical Anthropology* **117**: 810-818.

Martin, P., Garcia-Hirschfeld, J., Garcia, O., Gusmao, L., Garcia, P., Albarran, C., Sancho, M., Alonso, A., 2004. A Spanish population study of 17 Y-chromosome STR loci. *Forensic Science International* **139**: 231-235.

Mercieca, S., 2006. Marriage prospects in early modern Malta: the integration of Venetian subjects in an alien country. *Melita Historica* **14**: 303-324.

Mergen, H., Öner, R., Öner, C. 2004. Mitochondrial DNA sequence variation in the Anatolian peninsula (Turkey). *Journal of Genetics* **83**: 39-47.

Mifsud, M., 1995. *Loan verbs in Maltese: A descriptive and comparative study*. Brill. Leiden.

Muscat, J., 2000. The Tarxien ship graffiti revisited. *Melita Historica* **13**: 49-57.

Nei, M., 1972. Genetic distance between populations. *American Naturalist* **106**: 283-292.

Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583-590.

Novelletto, A., 2007. Y chromosome variation in Europe: Continental and local processes in the formation of the extant gene pool. *Annals of Human Biology* **34**: 139-172.

Ottoni, C., Martinez-Labarga, C., Vitelli, L., Scano, G., Fabrini, E., Contini, I., Biondi, G., Rickards, O., 2009. Human mitochondrial DNA variation in Southern Italy. *Annals of Human Genetics* **36**: 785-811.

Pääbo, S., 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences USA* **86**: 1939-1943.

Pääbo, S., Poinar, H., Serre, D., 2004. Genetic analyses from ancient DNA. *Annual Review of Genetics* **38**: 645-79.

Parson, W., Dür, A., 2007. EMPOP-A forensic mtDNA database. *Forensic Science International: Genetics*. **1**: 88-92.

Peakall, R., Smouse, P.E., 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology* Notes **6**: 288-295.

Pereira, L., Prata, M. J., Amorim, A., 2000. Diversity of mtDNA lineages in Portugal: Not a genetic edge of European variation. *Annals of Human Genetics* **64**: 491-506.

Poinar, H., Kuch, M., McDonald, G., Martin, P., Pääbo, S., 2003. Nuclear gene sequences from a late Pleistocene sloth coprolite. *Current Biology* **13**: 1150-1152.

Preciuttini, S., Caglia, A., Alu, M., Asmundi, A., Buscemi, L., Caenazzo, L., Carnevali, E., Carra, E., De Battisti, Z., De Stefano, F., Domenici, R., Piccinini, A., Resta, N., Ricci, U., Pascali, V. L., 2001. Y-chromosome haplotypes in Italy: the GEFI collaborative database. *Forensic Science International* **122**: 184-188.

Plaza, S., Calafell, F., Helal, A., Bouzerna, N., Lefranc, G., Bertranpetit, J., Comas, D., 2003. Joining the Pillars of Hercules: mtDNA sequences show multidirectional gene flow in the Western Mediterranean. *Annals of Human Genetics* **67**: 312-328.

Presciuttini, S., Cerri, N., Turrina, S., Pennato, B., Alu, M., Asmundo, A., Barbaro, A., Boschi, I., Buscemi, L., Caenazzo, L., Carnevali, E., De Leo, D., Di Nunno, C., Domenici, R., Maniscalco, M., Peloso, G., Pelotti, S., Piccinini, A., Podini, D., Ricci, U., Robino, C., Saravo, L., Verzeletti, A., Venturi, M., Tagliabracci, A., 2006. Validation of a large Italian database of 15 STR loci. *Forensic Science International* **156**: 266-268.

QiaGen, 2006. *QIAquick® Spin Handbook*. QiaGen. Germany.

Redd, A. J., Agellon, A. B., Kearney, V. A., Contreras, V. A., Karafet, T., Park, H., de Knijff, P., Butler, J. M., and Hammer, M. F., 2002. Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Science International* **130**: 97-111.

Richards, M., Côrte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H. J., Sykes, B., 1996. Paleolithic and Neolithic lineage in the European mitochondrial gene pool. *American Journal of Human Genetics* **59**: 185-203.

Richards M., Macaulay, V., Bandelt, H.J., 1998. Phylogeography of mitochondrial DNA in Western Europe. *Annual Review of Human Genetics* **62**: 241-260.

Richards, M., Macaulay, V., Hickey, E., 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *American Journal of Human Genetics* **67**: 1251-1276.

Richards, M., Macaulay, V., 2001. The mitochondrial gene tree comes of age. *American Journal of Human Genetics* **68**: 1315-1320.

Richards, M., Macaulay, V., Torroni, A., Bandelt, H-J., 2002. In search of geographic patterns in European mitochondrial DNA. *American Journal of Human Genetics* **71**: 1168-1174.

Rickards, O., Martinez-Labarga, C., Scano, G., De Stefano, G. F., Biondi, G., Capaci, M., Walter, H., 1998. Genetic history of the population of Sicily. *Human Biology* **70**: 699-714.

Robino, C., Inturri, S., Gino, S., Torre, C., Di Gaetano, C., Crobu, F., Tomano, V., Matullo, G., Piazza, A., 2006. Announcement of population data, Y-chromosomal STR haplotypes in Sicily. *Forensic Science International* **159**: 235-240.

Roewer, L. and Epplen, J.T., 1992. Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. *Forensic Science International* **53**: 163-171.

Romano, V., Cali, F., Ragalmuto, A., 2003. Autosomal microsatellite and mtDNA genetic analysis in Sicily. *Annals of Human Genetics* **67**: 42-53.

Rossi, E., Rolf, B., Schurenkamp, M., Brinkmann, B., 1999. Y-chromosome STR haplotypes in an Italian populations sample. **112**: 78-81.

Ruitberg, C. M. and Butler, J. M., 2000. New primer sets for Y chromosome and CODIS STR loci. (Poster presented at the Eleventh International Symposium on Human Identification, Biloxi, MS).

Rűstemoglu, A., Akay, G., Karabulut, H., Kadikiran, A., Bőkesoy, I., 2011. An investigation of 10 Y-STR loci and the detection of specific haplotype frequencies in Turkish population. *American Journal of Molecular Biology* **1**: 114-121.

Saitou, N., Imanishi, T., 1989. Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Molecular Biology and Evolution* **6**: 514-525.

Saunier, J. L., Irwin, J. A., Strouss, K. M., Ragab, H., Sturk, K. A., Parsons, T. J., 2009. Mitochondrial control region sequences from an Egyptian population sample. *Forensic Science International* **3**: 97-103.

Satoh, M., Kuroiwa, T., 1991. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Experimental Cell Research* **196**: 137-140.

Schoske, R., Vallone, P. M., Kline, M. C., Redman, J. W., and Butler, J. M., 2004. High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. *Forensic Science International* **139**: 107-121.

Schuelke, M., 2000. An economic method for the fluorescent labelling of PCR fragments. *Nature Biotechnology* **18**: 233-234.

Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, Moral P, Latini V, Varesi L, Memmi MM, Romano V, De Leo G, Gennarelli M, Jaruzelska J, Villems R, Parik J, Macaulay V, Torroni A., 2001. Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Human immunology* **62**: 871-884.

Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, P. J., Zhivotovsky, L. A., King, R., Torroni, A., Cavalli-Sforza, L. L., Underhill, P. A., Santachiara-Benerecetti, A. S., 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the

neolithization of Europe and later migratory events in the Mediterranean area. *American Journal of Human Genetics* **74**: 1023-1034.

Smith, C. I., Chamberlain, A. T., Riley, M. S., 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution* **45**: 203-217.

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A., Salas, A., Oppenheimer, S., Macaulay, V., Richards, M. B., 2009. Correcting for purifying selection: An improved human mitochondrial molecular clock. *The American Journal of Human Genetics* **84**: 740-759.

Stoneking, M., 2000. Hypervariable sites in the mtDNA control region are mutational hotspots. *American Journal of Human Genetics* **67**: 1029-1032.

Stoneking, M., Krause, J., 2011. Learning about human population history from ancient and modern genomes. *Nature Reviews Genetics* **12**: 603-614.

Sykes, B., Irven, C., 2000. Surnames and the Y chromosome. *American Journal of Human Genetics* **66**: 1417-1419.

Torroni, A., Lott, M. T., Cabell, M. F., Chen, Y., Lavergne, L., Wallace, D. C., 1994. MtDNA and the Origin of Caucasians: Identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-Loop region. *American Journal of Human Genetics* **55**: 760-776.

Troy, C. S., Machugh, D. E., Bailey, J. F. 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**: 1088-1091.

Trump, D. H., 2002. *Malta prehistory and temples*. Midsea Books Ltd. Malta.

Turchi, C., Buscemi, L., Previderè, C., Grignani, P., Brandstätter, A., Achilli, A., Parson, W., Tagliabracci, A. Ge.F.I. Group., 2008. Italian mitochondrial DNA database: results of a collaborative exercise and proficiency testing. *International Journal of Legal medicine* **122**: 199-204.

Turchi, C., Buscemi, L., Giacchino, E., Onofri, V., Fendt, L., Parson, W., Tagliabracci, A., 2009. Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: An enrichment of forensic mtDNA databases with Northern Africa data. *Forensic Science International* **3**: 166-172.

Turrina S., Atzei R., De Leo D., 2006. Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay. *International Journal of Legal Medicine* **120**: 56-59.

Urquhart, A., Oldroyd, N.J., Kimpton, C.P. and Gill, P., 1995. Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. *BioTechniques* **18**: 116-121.

Van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutation* **30**: 386-394.

Vella, A. P., 1979. *Storja ta' Malta*. Edizzjoni Klabb Kotba Maltin. Malta.

Vella, C., Vella Osanich, R., 2008. *Preliminary site report Tal-Gardina* (unpublished). Malta.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., Wilson, S., 1991. African populations and the evolution of human mitochondrial DNA. *Science* **253**: 1503-1507.

Vona, G., Ghiani, M. E., Calo, C. M., 2001. Mitochondrial DNA sequence analysis in Sicily. *American Journal of Human Biology* **13**: 576-589.

Walsh, P. S., Metzeger, D. A., Higuchi, R., 1991. Chelex®100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**: 506-513.

Watson, E., Forster, P., Richards, M., Bandelt, H-J., 1997. Mitochondrial footprints of human expansions in Africa. *American Journal of Human Genetics* **61**: 691-704.

Wayne, R. K., Leonard, J. A., 1999. Full of Sound and Fury: The recent history of ancient DNA. *Annual Review of Ecology and Systematics* **30**: 457-477.

Wettinger, G., 1986. The Arabs in Malta. *Malta: Studies of its heritage and history*. Ed. Mid-Med Bank. Malta.

Wettinger, G., 1999. The origin of the 'Maltese' surnames. *Melita Historica* **12**: 333-344.

White, P. S., Tatum, O. L., Deaven, L. L., and Longmire, J. L., 1999. New, male-specific microsatellite markers from the human Y chromosome. *Genomics*. **57**: 433-437.

Willerslev, E., Cooper, A., 2005. Ancient DNA. *Proceedings of the Royal Society B* **272**: 3-16.

Younge, C. D., 1903. *The orations of M. Tullius Cicero*. George Bell and Sons. London.

Young Lee, H., Song I., Ha, E., Cho, S., Yang, W. I., Shin, K., 2008. mtDNAmanager: a Web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences. *BMC Bioinformatics* **9**:483

Zalloua, P. A., Xue, Y., Khalife, J., 2008. Y-Chromosomal Diversity in Lebanon is structured by recent historical events. *American Journal of Human Genetics* **82**: 873-882.

Zalloua, P. A., Platt, D. E., El Sibai, M., Khalife, J., Makhoul. N., Haber, M., Xue, Y., Izaabel, H., Bosch, E., Adams, S. M., Arroyo, E., López-Parra, A. M., Aler, M., Picornell, A., Ramon, M., Jobling, M.A., Comas, D., Bertranpetit, J., Wells, R.S., Tyler-Smith, C., 2008. Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean. *American Journal of Human Genetics* **83**: 633-642.

Appendix 1

National strategic plan for the conduct of analysis using ancient DNA from inhumations found in the Maltese archaeological context.

Background:

Human burials have always been seen as a treasure trove of information by archaeologists. Up to a few years ago, these findings have always been interpreted through classical osteological means in an attempt to get a glimpse into the identity of the person that was buried, and also about the customs and ideas of his culture. These techniques whilst very useful have also got their limitations. These limitations are evident in kinship analysis and when an attempt at determining the sex of pre-puberty individuals is made; another limitation is that the experience of the osteologist plays a very big factor in the correct interpretation of the bones.

Research to date:

In the last 20 years, the advent of ancient DNA (aDNA) analysis has been made possible through the polymerase chain reaction (PCR). aDNA has been a vital tool for researchers when trying to garner as much information as possible from bone remains found in archaeological sites. Although plagued with problems of contamination, aDNA has been used in order to determine kinship between burials from the same site. Such a study was the much publicised case of the identification of the Romanov remains (Gill *et al.* 1994). aDNA has also been used in evolutionary biology in order to answer questions about the ancestors of modern man with research on Cro-Magnon and Neanderthal specimen determining that

Neanderthals did not contribute anything to the modern gene pool (Caramelli and Fox 2003). aDNA analysis has also branched out into the study of paleodisease which might not be evident on physical inspection of the bones, such as in the early stages of tuberculosis and leprosy. Some outrages claims which have damaged the reputation of aDNA research have also been made, such as the retrieval of DNA from Dinosaur specimen, and from insects trapped in amber which were tens of thousands years old. All such claims have been shown to be false and due to the inherent problem of contamination (Cooper and Poinar, 2000). In order to protect the integrity of their studies, aDNA researchers have agreed on guidelines that must be followed whenever one is working with aDNA. These guidelines, which will be listed further down in this document, ensure the least possibility of a false positive arising when working with aDNA.

Opportunities:

aDNA gives us an opportunity to study the genetics of a population throughout time; from when the matrilineal ancestors of the subject under study entered into Europe, up to how does the genetic profile of the individual compare to the modern day population that inhabits the area in which the burial was found. Such an analysis can also target specific questions which archaeologists cannot answer through traditional means, such as kinship analysis between burials, ethnicity, determination of the sex of infants and ambiguous adults, and also quantification of how many different people were present when a site is found with disarticulated skeletons in it, such as in secondary burial contexts.

Research Objectives:

The research objectives this plan aims to address are:

Identifying the potential of aDNA research being carried out in Malta, and determine the cut-off period where the thermal age of a site makes aDNA analysis from the said site not viable.

The setting up and upkeep of a database containing records of all aDNA research that is done on Maltese archaeological specimen, including all the retrieved sequences together with relevant information in a format that makes further research possible

Try to determine the genetic structure of the Maltese population during different times in Pre/History

Ensure that all work carried out conforms to accepted standard operating procedures for aDNA analysis.

The setting up of a scientific committee that acts as an advisory body to the Superintendence of Cultural Heritage in matters relating to the study of Human remains.

Methodology:

The preferred order of precedence by which the sites would be investigated would be as follows:

The most recently excavated sites would be investigated first: This is due to aDNA degrading at a faster rate once the inhumations have been unearthed during excavations. Thus in Malta, the sites which have been excavated after 1990 will be investigated first, since these are the sites which offer the major chance of extracting aDNA from them. If a special request is submitted for a research proposal which asks for samples from a specific site, the Superintendence will determine the importance of the research being attempted at the site and if sites with more pressing needs are identified, the Superintendence will encourage the researcher to submit a proposal concerning such sites with are deemed to be of greater importance.

The age of the site being excavated: The older a site is, the less the probability of extracting aDNA sequences from samples from the site. Thus sites which have a reasonable chance of yielding aDNA due to their age will be given precedence over older sites.

The climatic conditions of the site: DNA preservation is favoured in sites which are cool and dry. Thus if 2 sites of the same age are to be analysed. The inhumations of the site with the cool and dry conditions will be analysed first, as failure to retrieve aDNA from such a site means there will not be any surviving aDNA in the sister site with less suitable conditions.

The amount of people that came in contact with the sample: The greater the amount of people that come in contact with the sample, the greater the chance of contamination. Thus, sites in which few archaeologists worked on the human remains would be given preferential treatment. The archaeologists would also be contacted and asked if they consent to a saliva swab by which the genetic profile of the archaeologist would be mapped.

The order by which the extraction of DNA from inhumations from a site would be conducted is as follows:

The preservation of the bone would be the key factor. Samples which indicate poor biomolecular preservation would not be investigated

The inhumations archaeologists are most interested in, in order to solve archaeological questions from the site will be investigated first

The rest of the inhumations in good condition from the site will also be investigated, thus an attempt to build as complete as possible a genetic profile for the site would be made.

Quality control:

In order to reduce the risk of contaminating the samples, all work will be conducted according to the agreed on guidelines for the extraction and analysis of aDNA specimen set down by aDNA researchers (Cooper and Poinar, 2000). Some of these guidelines are:

Separate areas for the extraction of aDNA from the bones, amplification and post-PCR work: These areas should be physically separate from one another in order to ensure that no cross-contamination of the areas occur. No person who has entered the post PCR area is allowed in the other clean rooms for the rest of the day

The testing of all persons that came in contact with the samples from excavation until amplification.

Independent testing of important finds by a second laboratory

Funding:

aDNA analysis is time consuming and expensive, but can be very rewarding. A rough estimation of the cost of mtDNA haplotyping for a single sample from a site is in the region of €850. Y chromosome and autosomal STR DNA fingerprinting for a sample can go up to the region of €2000 for a sample, even if this figure is almost never reached due to the difficulty in amplifying Nuclear STR's from ancient specimen. Thus any project must ensure it has adequate funding for the proposed study.

Resources needed:

The most important factor in determining the successful outcome of such a project is the human element.

The osteologist who has worked on the site during excavations, and also afterwards on the bone samples, is invaluable in a project concerning aDNA. His role in the project is to indicate the bones which are best preserved, and also the samples which are of the most interest from the site, thus he also acts as a link between the site's principal investigators and the researchers.

The person conducting the study should have previous experience in sampling, DNA extraction, amplification and analysis. Ideally the person would have worked on aDNA before the Maltese project. The researcher must also be aware of recent innovations in the aDNA field, and also be able to adhere to the established practices used when dealing with aDNA to avoid contamination.

The person who maintains the National Database of aDNA sequences. This person must be skilled in archive keeping and computer literate in order to be able to store and retrieve different sequences as various needs arise.

When submitting a proposal the researcher must indicate: the methodology and type of analysis that is going to be conducted on the samples, his research objectives, who the people working with the remains are, which laboratories will be used for the study, and also a clear timeline for the project.

Laboratory/Materials:

Laboratory:

It is important to ensure that the Laboratory in which aDNA work is conducted conforms to the guidelines mentioned previously. Thus any laboratory working with aDNA must have a specialized clean room where DNA will be extracted from the bones. This clean room should be constantly irradiated with UV, in order to ensure cross-linking of any DNA fragments which might be present in the air. Another important requirement for such work is also another specialized clean room where the PCR reaction for aDNA specimen is set up. This clean room should also be constantly irradiated with UV light. Such a clean room is

vital because any modern DNA contamination will be amplified preferentially to the ancient one.

The above clean rooms should be physically separated from the Main Laboratory where post PCR work is conducted.

None of the above clean rooms should ever be used for the analysis of modern samples.

Materials/Supplies

All the reagents and supplies used in the process must be of the highest analytical standard, in order to ensure reliability of results. All equipment used should be calibrated and inspected in order to ensure that it performs as expected.

Timeline:

From previous experience working with aDNA samples, it is estimated that approximately 9 months are needed in order to process around 10-15 individual burials with pre-optimized primers. This timeline will be reduced if aDNA is found in the first attempts of amplification. Thus the researcher should submit a timeline of the proposed study and establish a date by which time the project will be finished, including the data analysis.

Expected results

The expected results will take the form of a series of DNA sequences, and analysis that is conducted by using these sequences. All data gathered from the research will be published in a series of papers in Scientific Journals, and also be included in the site report. This data is expected to include haplotyping of individual samples, kinship analysis of site burials, comparisons between populations of different sites and different timelines.

The researchers will also be expected to present the findings at an international conference in either poster format or a talk, in order to be subjected to the scrutiny of their peers.

Beneficiaries:

The Beneficiaries of projects adopted under this national plan will be:

The people of Malta: as such studies will strengthen the Maltese identity and the perception of roots in the island.

The Superintendence of Cultural Heritage, who will acquire a database of aDNA sequences which can be used in future studies about Maltese population genetics.

The Central Mediterranean region, which will have the Maltese projects as case studies, with which other countries and regions can assess if such work is feasible in their own country, with similar climatic conditions.

The funding bodies who will get prominent mentions and acknowledgments in media, scientific papers and in any posters and presentations given

The **researcher** who will gain academically and also further experience in analyzing aDNA remains.

Bibliography:

Caramelli, D., Laluez-Fox, C., Vemesi, C., Lari, M.,2003. Evidence for a genetic discontinuity between Neanderthals and 24,000-year-old anatomically modern Europeans. *Proceedings of the National Academy of Sciences*, **100**: 6593-6597.

Cooper, A., Poinar, H. N., 2000. Ancient DNA: Do it right or not at all. *Science,* **289**: 1139.

Gill, P., Ivanov P. L., Klimpton, C., Piercy, R., 1994. Identification of the remains of the Romanov family by DNA analysis. *Nature Genetics*, **6**: 130- 135

Appendix 2

MtDNA HVR1 polymorphisms of persons who came in contact with the aDNA samples.

Archaeologists, curators and laboratory members MtDNA HVR1 sequences: Possible Sources of contamination (No distinction was made between the different groups in order to increase anonymity).

Possible Contaminants	**Polymorphisms different to CRS**
Possible Contaminant 1	None
Possible Contaminant 2	16189d
Possible Contaminant 3	None
Possible Contaminant 4	None
Possible Contaminant 5	16125A
Possible Contaminant 6	16243C
Possible Contaminant 7	None
Possible Contaminant 8	16289C
Possible Contaminant 9	16148T 16210C 16246A
Possible Contaminant 10	16243C 16280G 16303G 16305C
Possible Contaminant 11	16243C 16270A
Possible Contaminant 12	16289C
Possible Contaminant 13	16168G 16246C 16263A 16300C 16306C
Possible Contaminant 14	None
Possible Contaminant 15	None
Possible Contaminant 16	None
Possible Contaminant 17	None
Possible Contaminant 18	None
Possible Contaminant 19	16341C

Possible Contaminant 20	16301T
Possible Contaminant 21	16138G 16189C 16242T
Possible Contaminant 22	None
Possible Contaminant 23	16169T 16223T 16278T 16294T 16309G 16390A
Possible Contaminant 24	16169T 16223T 16278T 16294T 16309G 16390A
Possible Contaminant 25	16126C 16153A 16294T

Positions were checked more than once when possible in order to determine any sequencing errors. In laboratory members where the researcher only obtained the mutation motif from the possible contaminant but not the sequencing chromatogram , rechecking sequencing quality was not possible.