UNIVERSITY OF BOLOGNA

Ph. D. Program in

"Biodiversity and Evolution"

Cycle XXIII

BIO/08 - Anthropology

# THE GENETIC SIGNATURE OF NEOLITHIC IN GREECE

PAOLO ANAGNOSTOU

Coordinator:

Prof. Barbara Mantovani

Supervisor:

Prof. Donata Luiselli

Academic Dissertation, 2011

# TABLE OF CONTENTS

# ABSTRACT

The Neolithic is characterized by the transition from a subsistence economy, based on hunting and gathering, to one based on food producing. This important change was paralleled by one of the most significant demographic increase in the recent history of European populations. The earliest Neolithic sites in Europe are located in Greece. However, the debate regarding the colonization route followed by the Middle-eastern farmers is still open. Based on archaeological, archaeobotanical, craniometric and genetic data, two main hypotheses have been proposed. The first implies the maritime colonization of North-eastern Peloponnesus from Crete, whereas the second points to an island hopping route that finally brought migrants to Central Greece. To test these hypotheses using a genetic approach, 206 samples were collected from the two Greek regions proposed as the arrival point of the two routes (Korinthian district and Euboea). Expectations for each hypothesis were compared with empirical observations based on the analysis of 60 SNPs and 26 microsatellite loci of Y-chromosome and mitochondrial DNA hypervariable region I. The analysis of Y-chromosome haplogroups revealed a strong genetic affinity of Euboea with Anatolian and Middle-eastern populations, while the Korinthia District population shows a closer affinity with Balkan populations. The inferences of the time at which population expansion started suggests an earlier evidence of usage of agriculture in Euboea. Moreover, the haplogroup J2a-M410, supposed to be associated with the spread of the Neolithic lifestyle from the Middle-east, was observed at higher frequency and variance in Euboea showing, for both these parameters, a decreasing gradient moving from this area. The time since expansion estimates for J2a-M410 was found to be compatible with the Neolithic transition in Greece and slightly older in Euboea. The analysis of mtDNA resulted less informative. However, a slightly higher genetic affinity of Euboea with Anatolian and Middle-eastern populations was confirmed. These results taken as a whole suggests that the most probable route followed by Middle-eastern or Anatolian populations during the colonization of Greece was the island hopping route that brought them in Central Greece. Further studies, using a genomic approach, could bypass the limitations of uniparental markers and draw a clearer picture of the Neolithic migrations into Greece.

# 1. THE NEOLITHIC IN GREECE AND EUROPE

The transition from hunting and gathering to agriculture is certainly one of the most significant events in human prehistory, representing a shift from foraging to farming, from food collection to food production, from wild to domestic, which sets the stage for most of the subsequent developments in human society (Price 2000, Ammerman & Cavalli-Sforza 1984). For this reason, the beginning of agriculture have been a subject of interest since the middle of the 19th century when Charles Darwin in 1869 treatise this subject in "Variation of plant and animals under domestication".

### 1.1 The origin of agriculture and birth of the first farming communities

This and the following paragraphs represent an overview of the knowledge, accumulated over the past years, regarding the spread and emergence of the Neolithic culture in Europe, which provides the necessary background to the genetic investigation. The literature on this subject is extremely vast heterogeneous, ranging from pure archaeology to sociology thus difficult to disentangle and summarize. Therefore, I have selected from the literature and combined concepts, ideas and conclusions which seem the most reliable and extensively used them in the first paragraphs of this introductory chapter.

The reasons why agriculture had been a successful way of life and had overcome the hunting and gathering lifestyle it is its unquestionably superior mode of production. Darwin for instance, believed that, knowledge was the determining factor for the beginning of agriculture. In fact, he suggested that once the knowledge that a seed planted in the ground would grow into a plant discovered, then food production would be the following step.

Several theories have been, in times, proposed in order to explain the reasons of the birth of agriculture and the success that this new subsistence economy had all over the world. Three of them are worthy to be recalled.

One of the most known theories is the "Oasis theory" proposed by Childe in (1928; 1934). According to this theory the end of the last Ice Age led to a great increased drought of large areas. This brought all animal, including humans, to concentrate in those river valleys where water was still available. This increased degree of proximity

between different species led to new relationships between them. In these highly favoured environments, where all varieties of plant food would grow, stubble from harvested crops would be a significant food source for grazing animals. The interests of people and animals therefore coincided, and agriculture was born.

On the other hand Cohen (1977; 1989; Cohen and Armelagos 1984) places emphasis on population increase as the source of the imbalance between resources and human numbers. He assess that the only factor which could possibly account for the irreversible and nearly uniform development of agriculture on a global scale was a growth in population levels beyond the point that hunting-gathering could support. While agriculture did not necessarily improve the overall diet or make it more reliable, it did provide a higher calorific output per unit of time invested in comparison with what hunting-gathering could do. At first, attempts were made to intensify resource use through such measures, but as population growth continued its increasing trend, these efforts proved insufficient. In the end the only remaining solution was the development of a food-producing economy.

The third theory briefly described here is that of MacNeish. This author, trying to pool together the above theories, developed a model, in which he produces a series of *necessary* and *sufficient* conditions for the emergence of groups of initial agriculturalists (MacNeish 1992). These necessary conditions are:

(1) An ecologically highly diverse environment

(2) The existence of potentially domesticable plants in one or more ecozones

(3) The exploitation of a variety of resources which cannot all be reached from a single base

(4) A natural seasonal cycle, with a harsh season when few resources are available

(5) A gradual rise in population

While his *sufficient* conditions are:

(1) A change in the environment which reduces available resources and, in particular, makes the harsh season worse

(2) An increase in the degree of sedentism leading to further population pressure

(3) An increase in the structuring of food provision, with a wider range of resources exploited and the use of storage

(4) A change in the ecosystem and/or in the genetic makeup of some of the seeds usually collected, reducing the energy spent in gathering each seed.

If all or at least some of these conditions were present, permanent villages with an increasing population will rise and thus making the increase of food production a necessity, hence agriculture.

This conditions proposed by MacNeish were present in the Fertile Crescent around the 12th millennia B.P. On the basis of archaeological, archaeobotanical and paleo-anthropological reports (Van Andel & Runnels 1995; Thorpe 1996; Colledge et al., 2004; Pinhasi & Pluciennik 2004 and Pinhasi et al., 2005) the origin of agriculture it is to be conducted among the Natufian populations that left evidence of collection or harvesting of wild cereals, usage of pounding tools such as mortars and pestles and a more intense use of cereals in their diet. Moreover, evidence of permanent settlements is reported for the Hayonim Cave, located in the upper Galilee in Israel. The presence of continued use of the same site by members of a single lineage is indicated by the presence, in successive layers of the burials, of individuals with a congenital absence of the third molar (Smith *et al.* 1984); this condition occurs among this particular group more frequently than in the Natufian population as a whole. The further development of the farming technology, among the Natufians, transformed their societies to a state that we now call Neolithic. A more recent investigation based on radiocarbon dates of 735 archaeological sites all over the Middle-east, Anatolia and Europe (Pinhasi et al., 2005) places the putative origin of agricultural practices in the southern Levant and the southern Mesopotamia, more specifically in the area that comprise the Neolithic sites of Jericho (Jordan), Aswad (Syria) and Abu-Hureyra (Syria).

From there the "Neolithic package" composed by: permanent villages of rectangular houses, religious objects and structures, domesticated plants (einkorn, wheat, millet, spelt and burly ) and animals (pigs, sheep and goats) as well as pottery and ground tools had spread in Anatolia first and then in Europe, reaching the Aegean coast sometime around the 9th millennium B.P. (Demoule & Perlès 1993).

### 1.2 The spread of Neolithic in Europe: an archaeological overview

The expansion of agriculture in Europe took place relatively quickly. In less than 3000 years the agriculture technology reached the British islands and Scandinavia (Figure 1.2a), much quicker than the spread of farming in Southeast Asia (Thorpe 1996).

The advent of Neolithic lifestyle in Europe can be divided into 4 main episodes: the spread in the Aegean, the spread in the Balkan peninsula, the spread in the Western

Mediterranean (maritime colonization of Southern Italy and the subsequently spread to costal Mediterranean France and Spain) and, finally, the spread in Central Europe (the Linear BandKeramik culture) and the Northern Europe.

The first episode is characterized by the maritime colonization of Crete and the spread of farming technology into mainland Greece. This event will be described later in this chapter in more details.

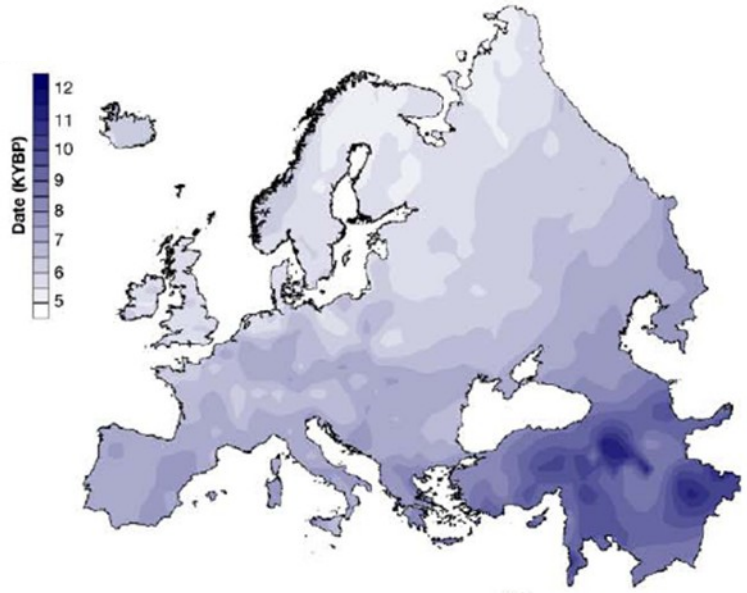



**Figure 1.2a**. Isochron map representing dates of early Neolithic sites in Europe based on data of Pinhasi et al., 2005 (reviewed in Balaresque et al., 2009).

### 1.2.1 The Neolithic in Western Mediterranean

The arm of the early Neolithic, spreading from the eastern to the western Mediterranean, sometime around 8Kya, is characterized by the appearance of stylistically uniform ceramics (Cardial pottery) (Figure 1.2.1a), domesticated plants and animals, and the exchange of items such as obsidian. The adoption of these features appears to have been taken place in two stages. The first stage carried from the east along the Dalmatian coast to Italy and southern France during the first half of the 8th millennium B.P. (Donahue 1992, Binder 2000). The second stage of this expansion covered the area from southern France around the coast of Spain to the Atlantic coast of Portugal in the second half of the millennium (Zilhao 2000). The sporadic and costal distribution of early Neolithic sites in these areas has been considered as the result of a

maritime colonization rather than inland colonization or a cultural diffusion of agriculture. These routes have been mainly deduced by the spread of the Cardial pottery.

**Figure 1.2.1a**. Example of Iberian Cardial pottery. Small pot with impressed cardial decoration. Museu Arqueologic Municipal Camil Visedo Molto (Alcoi, Alicante/Espana).

The appearance of the Cardial culture varies from east to west (Whittle 1996). In Dalmatians the Cardial pottery dated in the first years of the 8$^{th}$ millennium B.P. . In this area the complete "Neolithic package" appeared only in costal sites, not in the interior. In the same times, in southern Italy the Cardial culture is characterized by the presence of large farming communities dating around 8Kya (Donahue 1992, Pluciennik 1998) which also exhibits, the whole "Neolithic package". On the other hand, in the Po basin, in North-eastern Italy, the "Neolithic package" was adopted much latter (7250-6750 B.P.) and not as a whole (mainly by the adoption of ceramics) (Bagolini 1990, Lanzinger et al., 2000). This evidence point to a more probable cultural diffusion of farming technologies in Northern Italy, while the southern areas were, most probably, colonized by farmers.

The Cardial tradition quickly spread along the Tyrrhenian coast between 6-5,6 Kya. The Mesolithic and Neolithic sites in southern France and north-western Italy have a restricted coastal distribution, suggesting the spread of a colonizing Neolithic population along the coast (Price 2000).

The second stage of expansion, dating around 7,4Kya, saw the appearance of Cardial materials in the west Mediterranean region of France and the coasts of the Iberian peninsula. Evidence from the Aude valley in France suggests the local evolution of projectile points and lithic industry during the early Neolithic (Price 2000). Moreover, evidence of a local adoption of agriculture, almost 1000 years later than France, is also reported for the Cantabrian Spain and Portugal.

In conclusion, the very rapid spread of Cardial culture could suggest the prior existence of communication routes. The question regarding who carried domesticated plant and animals and ceramics along these routes is still open. The archaeological evidence indicates the spreading mechanism was very complex, including both colonization and local adoption of farming (Price 2000).

### 1.2.2 The Neolithic in the Balkan peninsula

Less than 1000 years after the agriculture's first appearance in Greece, in which was already well established, the first Neolithic settlements started to appear in the interior of South-eastern Europe. The radiocarbon dates for the early Neolithic sites in the Balkans are 8,6-7,8 Kya and 8,5-7,2 Kya for the southern the northern areas respectively (Chapman & Dolukhanov 1993; Tringham 2000). The most accepted theory on the spread of Neolithic into the Balkans involves the expansion of farmers from the plains of central and northern mainland Greece moving along the natural corridors of the major river valleys. In fact, areas like the Vardar-Morava corridor, the Maritsa basin and the middle and lower Danube basin saw the simultaneous arrival of early Neolithic culture. With just a few exceptions all early Neolithic sites in this area exhibits strong similarity, as it concerns their structure, with the Middle-eastern ones. This supports a model of colonization by immigrant farming populations, even if this point is still much debated. In fact, several scholars (Radavanovic 2000; Bogucki 2009) believe that the evidence for migrating farmers in the northern Balkan peninsula are not very informative since deep tell deposits do not easily reveal their lowest levels. Furthermore, the substantial variation in faunal assemblages at early Neolithic sites in Balkans mainly indicates local adoption rather than migration of farmers (Greenfield 1993).

### 1.2.3 The Neolithic in Central Europe: the Linear BandKeramik Culture (LDK)

The first Neolithic communities in central Europe have been always thought to be BandKeramik groups, based on the homogeneous patterns seen for ceramic shape and design (Figure 1.2.3a).

**Figure 1.2.3a**. Example of Linear Bandkeramik pottery. From the collection of the Gallo-Romeins Museum Kielenstraat.

This culture originated around 7,5Kya in villages along the middle Danube in eastern Hungary. The expansion of BandKeramik culture, as for the Cardial one, was extremely rapid. Within a time span of less than 200 years, around 7,5Kya, small farming settlements appeared in a large area which goes from today's Belgium and northern France to central Europe and Ukraine. Settlements distribution was not continuous; they seem to be concentrated in well-watered valleys. These communities, in addition to BandKeramik pottery, shared also other features such as architecture, artefacts, burials and settlements plans. All this features has been initially taken as a clear example of colonization by farming groups (Whittle 1997). Although, new archaeological evidence began to crumble these initial believes, suggesting a greater role for Late Mesolithic inhabitants in the development of LBK culture (Groneborn 1999). Groneborn point out that the initial phase of the LBK in central Europe is characterized by significant

heterogeneity among assemblages, suggesting local changes rather than homogeneous colonization. This view suggests that the LBK was just another example of hunters-gatherers becoming farmers.

Along the Atlantic and Baltic coasts, at the margins of the LBK dispersal, the presence of Mesolithic groups somehow blocked the spread of agriculture for almost 1500 years. In fact, the archaeology of early farming in Scandinavia and the British islands is extremely different from the LBK (Price 2000). For both regions the first evidence for the Neolithic appears around 6Kya with Funnel Beaker pottery, large tombs and bog sacrifices (Price 2000). Moreover, the difficulties in distinguishing between Mesolithic and Neolithic assemblages, in both Scandinavia and British islands, support an interpretation involving local adoption of agriculture.

### 1.3 A genetic overview on the spread of Neolithic in Europe

Since the pioneer work of Menozzi, Piazza and Cavalli-Sforza using classical markers, genetic data, prevalently uniparental markers, have been used to investigate past human history. Of course the Neolithic transition was not an exception.

The first studies performed for Y-chromosome (Rosser et al., 2000; Semino et al., 2004) produced, for certain haplogroups (J2-M172 and E1b-M78), gradients similar to the classical demic diffusion hypothesis, supporting the "Wave-of-advance" model. Subsequent studies have restricted the list of haplogroups that could be taken as markers of the Neolithic transition in Europe, pointing to the J2-M172 and G2a-P15 lineages (Cinnioglu et al., 2004). In the last few years several studies approached the issue of the Neolithic impact in Europe. Sengupta and colleagues in 2006 and subsequently, King et al. (2008) suggested the haplogroups J2a-M410 as the main haplogroup linked to the Neolithic spread from the Middle-east. Furthermore, King proposed that the G2a-P15 lineage too could be a marker of such migration, especially in Europe.

One recent paper (Balaresque et al., 2009) has investigated the Neolithic transition in Europe, by focusing on the main western European Y chromosome haplogroup R1b-M269. This lineage had received little recent attention in this context, since has been always considered to be Palaeolithic in origin (Semino et al., 2000; Rosser et al., 2000; Wilson et al., 2001) and therefore improbable to have been carried into Europe by the migrating farmers. Balaresque and colleagues used several Y chromosomes within

haplogroup R-M269 to show that the associated microsatellite diversity distribution cline goes to the opposite direction compared to the strong cline, observed for the frequency distribution, from high frequencies in the west which gradually decreases eastward (Figure 1.3a).

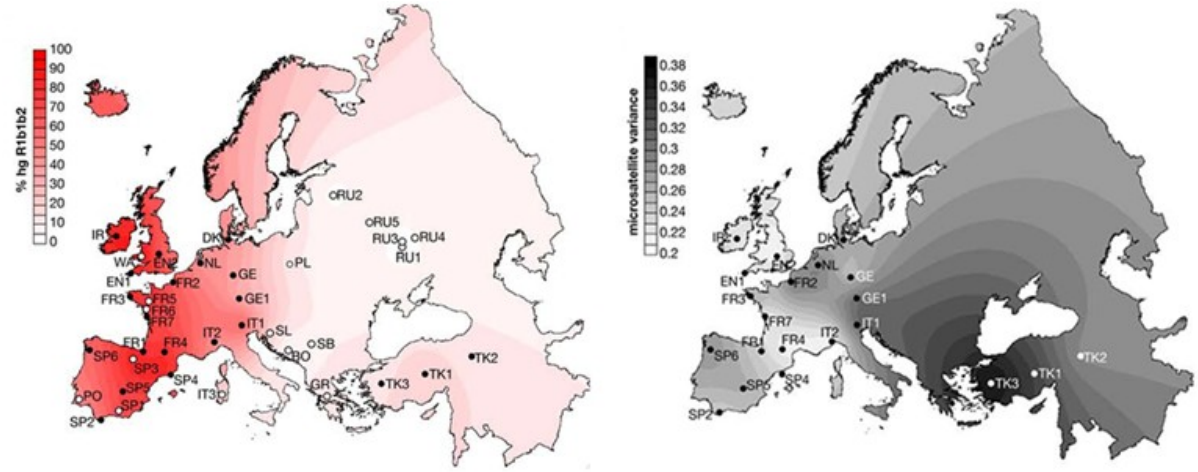



**Figure 1.3a**. Geographical distribution of haplogroup frequency of R1b-M269 (in the left) and mean microsatellite variance (in the right). From Balaresque et al., 2009.

They posited that this correlation could be explained by a more recent dispersal of this lineage from the Anatolia coinciding with the Neolithic agricultural transition in Europe. The dating of the lineage to approximately 6Kya in various populations seems also to support this idea. Moreover, this result, according to the authors, indicates that the great majority of the Y chromosomes of Europeans have their origins in the Neolithic expansion, supporting thus the "wave-of-advance" model.

On the contrary, Morelli and colleagues compared Y chromosome variation in Sardinia and Anatolia in male sample spread among several haplogroups. Their major result suggests that the specific motif defined by two microsatellite loci alleles they found, could separate haplotypes within a network of the R-M269 haplogroup into two well defined branches, geographically located either in Anatolia or in Sardinia (Morelli et al., 2010). Dating of the two branches indicates that the diffusion of agriculture from the Middle-east did not involve the significant movement of R-M269 chromosomes, and that haplogroup R-M269 was already present in Sardinia (and thus in Europe) prior to the Neolithic expansion.

An even more recent paper, Myres et al. (2010) described several new SNP mutations within the R-M269 lineage which show strong geographic structuring. They detected a main European specific clade, characterised the M412 and L11 mutations,

which is clinal from high frequencies (>70%) in Western Europe, decreasing eastward, in accordance with Morelli et al., 2010. Myres et al show that the distribution of several downstream SNPs exhibit remarkable frequency patterns that appear to spread from different areas of highly frequencies. Moreover, they estimated coalescence times for the S116 sub-branch in different populations in Europe suggesting a association with the spread of LBK culture.

In contrast, other studies underline the importance of local adoption in the process of diffusion of agricultural technologies. Capelli et al., in 2007, analysing the Y-chromosome diversity throughout the Italian peninsula, proposes a model of initial demic diffusion followed by a certain degree of local adoption, which increases heading northward. Interestingly, in the case of Italy, the Apennine seems to have acted as a barrier since the genetic contribution of Middle-eastern farmers reaches higher latitudes in the Adriatic side of Italy rather than the Tyrrhenian one. The importance of cultural diffusion is emphasized in Battaglia et al., 2009. These authors analyzing subjects from several Balkan populations assess that the internal diversity of three Y-chromosome lineages (I-M423, E-V13 and J-M241) can provide evidence to distinguish between the Holocene Mesolithic forager and subsequent Neolithic expansions from the Middle-east. In particular, whereas the Balkan microsatellite variation associated to J-M241 correlates with the Neolithic period, those related to E-V13 and I-M423 Balkan Y chromosomes are consistent with a late Mesolithic time frame. Thus, the Balkan Mesolithic people were thought to be the first to adopt farming, introduced by a group of migrating farmers from the Middle-east. These first locally converted farmers became the main agents of the spread of this economy in the Adriatic and the transmission of the "Neolithic package" to other adjacent Mesolithic populations.

The Neolithic contribution to the mitochondrial European gene pool is supposed to be very small (15%) (Soares et al., 2010). Although, the increasing use of ancient DNA data from Neolithic farmers can offer a direct view on the genetic past. To this regard two recent studies have addressed the issue and proposed different scenarios. The first one, through the analysis of mtDNA types from central and northern European post-LGM hunter-gatherer skeletal remains and homologous mtDNA sequences from early farmers(Bramanti et al., 2009), suggests that the observed lack of genetic continuity between the farmers and contemporary Europeans, points towards a prevalent pre-Neolithic contribution to the current Central European gene pool. On the other hand, a

second study conducted by Haak and colleagues (2010), concludes that the LBK population shared an affinity with the modern-day Middle-east and Anatolia, supporting a major genetic input from this area during the advent of farming in Europe. This result is based on the comparison of a large dataset of mtDNA types from skeletal remains of Central European LBK farmers, an ancient mtDNA dataset of hunters-gatherers and a wide dataset of extant West Eurasian populations. In the LBK dataset the U4 and U5 haplogroups, which together composed the 80% of the hunters-gatherers diversity, were virtually absent, pointing to a lack of temporal continuity between Mesolithic and Neolithic populations. Moreover, the mtDNA haplogroup composition of the LBK would suggest that the input of Neolithic farming cultures (LBK) to modern European genetic variation was much higher than that of Mesolithic populations. The LBK dataset shared a higher number of "Neolithic" informative haplotypes with Iranians and other Middle-eastern populations, suggesting a higher affinity with these latter populations. All these results have been taken from Haak et al. as evidence supporting a demic diffusion from the Middle-east. The genetic distance map generated to visualize similarity/distance of the LBK dataset to all modern populations, according to the authors, could suggest the geographic route of the dispersal. This route starting from eastern Anatolia, goes westward across the Balkans, and then northwards into Central Europe (Figure 1.3b)

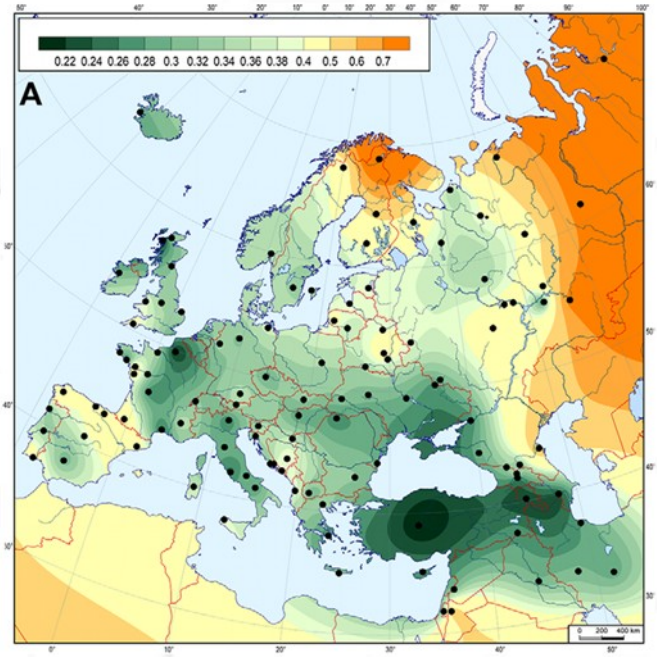



**Figure 1.3b**. Mapped genetic distances between Eurasian populations and the LBK dataset. short distances (greatest similarity) are marked by dark green and long distances (greatest dissimilarity) by orange, with fainter colours in between the extremes. From Haak et al., 2010.

### 1.4 Cultural or demic diffusion? An open question

Both archaeological and genetic studies regarding the transition to farming in Europe have been extensive and extremely meticulous. Although, one question remains, still in our days, unanswered. Did the farming technology spread by cultural diffusion or was brought along with migrants from the Middle-east? This issue was first stressed in 1958 by Gordon Childe in his book "The dawn of European civilization". He proposed that the first Neolithic crops and domesticated animals did not reach Europe by "*means of trade or exchange, but by means of migration or the colonization of farmers and shepherds from the Middle-East*". This idea was further stressed by Ammerman and Cavalli-Sforza in the early seventy's (1973). They produced a mathematical model of local migration, out from an area of population increase, in a wave pattern which radiates outward at a constant rate. This model called "wave-of-advance", borrowed from the field of population biology, proposes that population growth at the periphery of the farmers range, together with local migratory activity, would produce a population range expansion that moves outwards in all directions and advances at a relatively steady rate. They also predicted that the mixing of Neolithic and Mesolithic populations would lead to genetic gradients with extreme gene frequencies in the areas with the oldest Neolithic sites. This prediction turned out to be confirmed by the analysis of classical genetic markers (Menozzi et al., 1978). The first principal component of the classical polymorphisms shows a geographic cline, from the south-east to the north-west of Europe, as expected under the hypothesis of demic diffusion and the interaction of Neolithic and Mesolithic populations. The cline highlighted by this analysis shows a strong correlation between genetic and archaeological distances, and this correlation was considered as a support to the hypothesis of demic diffusion.

The "wave-of-advance" model has been hardly criticised by several archaeologists, especially those in favour of a substantial Mesolithic contribution to the agricultural transition. At a general level, they argued the ability of farmers to overcome hunter-gatherers. They assess that early agricultural communities would have been extremely vulnerable to competition from the existing populations: their permanent settlements could easily be attacked, their crops destroyed and livestock killed or taken (Dennel

1985). Moreover, they poses a series points of disagreement: first, in many areas of western and northern Europe there is a strong degree of continuity in some aspects of material culture (such as lithics) and a preservation of important symbols (for example bears, water birds and fish); second, the changes in physical anthropological aspects could result from shifts in diet rather than the replacement of the native population by migrant farmers, and that the genetic patterns saw by Ammerman and Cavalli-Sforza and others (Sokal *et al.* 1991) could not be related to the Neolithic; third, the process of the adoption of agriculture is much more slow than the "wave-of-advance" model would predict; fourth, this model underestimates the potential of gatherer-hunter population to change and innovate (Zvelebil and Zvelebil 1988).

Since the advent of genetic data, several studies tried to answer the question of the modality of the spread of farming in Europe. The first study, as reported earlier, was conducted on classical markers by Menozzi, Piazza and Cavalli-Sforza in 1978. The "wave-of-advance" model, proposed in this study, predicts a high genetic input of migrant populations into the European gene pool. Some mitochondrial-DNA studies suggest that the contribution of Middle-eastern farmers to the European gene pool is about 20% (Richards et al., 2000). A similar percentage (22%) is suggested by a Y-chromosome study carried out by Semino et al. (2000). A more recent study that makes use of mitochondrial-DNA, Y-chromosome DNA, and other autosomal markers (Dupanloup et al., 2004) finds that the Neolithic contribution is much higher than 20%, and decreases from east to west, as expected under the "wave-of-advance" model, even if local fluctuations are present. Thus, many genetic studies seem to support the idea of demic diffusion at some level, but there is still a lack of agreement regarding the percentage of the contribution of early middle-eastern farmers to the European gene pool. In the last few years, as we saw in the previous paragraph, a number genetic studies, for both Y-chromosome and mtDNA (Battaglia et al., 2009; Bramanti et al., 2009; Myres et al., 2010; Soares et al., 2010; Morelli et al., 2010) tend to support a model of local cultural adoption of agriculture. On the other hand, several other studies tend to confirm the "Wave-of-advance" model (Balaresque et al., 2009; Haak et al., 2010).

The conclusion that emerges from all these studies is that the spread of agriculture in Europe was an extremely complex process that cannot be reduced as done by the two majorly accepted models. Many other models which consider both demic and cultural diffusion have been, in times, proposed. The most promising is the "pioneer model"

which postulates that there was an initial, small scale migration of farmers from the Near East to certain regions of Europe. These small groups have, probably, undergone localized demographic expansions due to social advantages. The subsequent spread of farming technologies throughout the rest of Europe was then carried out by Mesolithic Europeans who adopted this new technology through trade and cultural interaction (Zvelebil 2000).

### 1.5 The Neolithic in Greece

As archaeological reports reveal, the first European farming communities are found in Greece. The Mesolithic (10,7-9 Kya) landscape of Greece it is, unfortunately, poorly known (Perlès 2001) and just a few sites are recorded. This few sites known to date, concentrate in two main regions: north-east Attica and the Argolide, in eastern Greece, Corfu, the coastal plains of the Acheron and the Preveza region in north-western Greece (Figure 1.5a).


**Figure 1.5a**. Sites and locations dating to the Mesolithic period in Greece. From Perles 2001

The significance of this scarcity has led to opposing views on the origin of Neolithic in Greece. The demography of the Mesolithic background contributed to local innovations and to the role that local groups could have played in the constitution of the first farming societies (Nandris 1977). Many authors have considered the scarcity of Mesolithic sites, in contrast to the large number of Early Neolithic settlements, as good evidence for a demic diffusion of Neolithic groups into south-eastern Europe (van Andel and Runnels 1995). On the other hand, as suggested by Chapman (1991) this scarcity of Mesolithic sites could be an artefact due to the lack of intense research or the destruction of the sites. Therefore, if the Mesolithic sample is biased by geological factors or by lack of research, then no valid comparison can be drawn between the Mesolithic and the Neolithic. Consequently no conclusion can be reached regarding the origins of the latter (Andreou et al., 1996). Although, the lack of element related to local Mesolithic traditions in the first farming settlements seems to strongly suggest that farming was introduced in Greece by immigrants.

The peopling picture of Greece changes radically with the advent of the Neolithic. The extremely high number of early Neolithic sites seems in sharp contrast with the scarcity of Mesolithic ones (Figure 1.5b) (Perlès 2001).

The distribution of the nearly 250 early Neolithic sites could lead to the conclusion that a progressive and regular expansion occurred. A closer examination of the site's distribution, however, suggests that this does not hold true. In fact, not all regions were settled, the density of settlement varied widely. The Early Neolithic is well represented, in Thessaly in Euboea island and Attica and in the North-eastern regions of the Peloponnesus. On the contrary, it has not, so far, been documented in central and eastern Macedonia, in Thrace, in the western central Greece and in the Sporades and Cycladic islands. Probably early Neolithic farmers had well defined criteria in the choice of regions in which they chose to settle. In contrast with Mesolithic hunters-gatherers, which preferred to settle in hills and mountains, the early Neolithic farmers preferred the alluvional plains and basins. In fact, in Greece, from the nearly 250 early Neolithic sites only 3 hill-sites were definitely occupied during this period. However, not every alluvional basin presents the same density of permanent settlements. This is another important contrast which opposes eastern and western Greece. As it is clearly shown by Figure 1.5b, the majority of the sites concentrate on the eastern half of Greece.

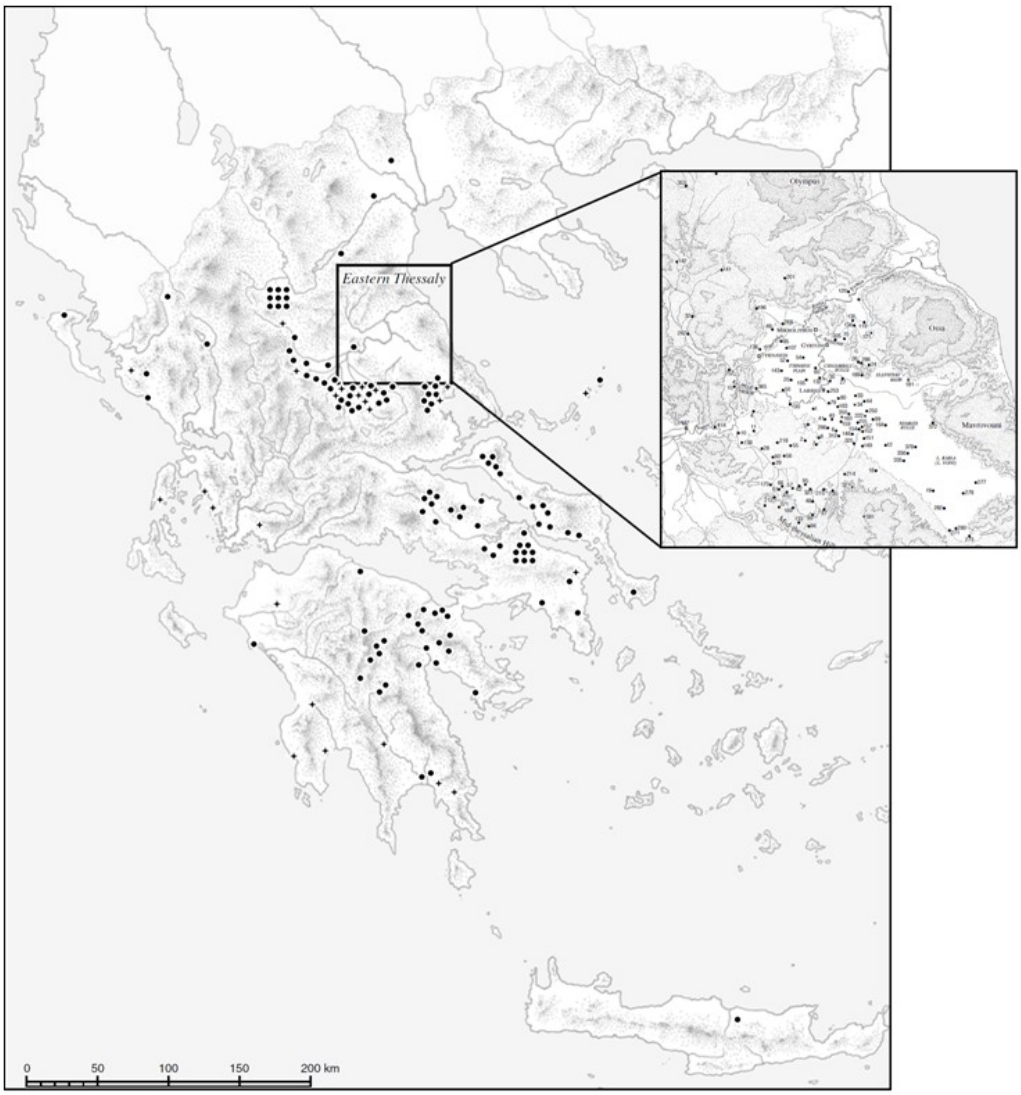



**Figure 1.5b**. Sites and locations dating to the early Neolithic in Greece. From Perlès (2001).

The east–west distribution of Early Neolithic sites closely follows hydrometric curves (Polunin 1987). The vast majority of known early Neolithic sites are located in the driest part of Greece, with a mean annual precipitation of about 600 mm or less. This could be, apparently, a paradox; however, the regions exploited present the natural conditions closest to those of the Middle East, which minimizes the problems of species adaptation to new environmental conditions and in particular to higher precipitations, a major problem for cereals (Hillman and Davies 1990).

Even within eastern Greece the density of sites shows important regional variations, with higher density in the North and decreasing southwards. This decreasing gradient approximately parallels the gradients of increasing mean temperatures, decreasing annual rainfall and, consequently, increasing inter-annual fluctuations (Perlès 2001). According to van Andel and Runnels (1987), early agricultural settlements in southern

Greece were constrained by the necessary presence of permanent springs. This hypothesis was confirmed by a detailed study of settlement patterns in the Argolide and Korinthia which were located in areas with abundant groundwater, near spring-watered meadows, perennial streams and lakes, reflecting the choice of a homogeneous type of landscape with well-watered alluvial soils of high potential for arable agriculture (Johnson 1996).

### 1.5.1 The Neolithic migration into Greece

It is now unarguable that farmers of Middle-eastern origin migrated into Greece bringing along the full "Neolithic package". Moreover, the route followed by these migrants is still a matter of debate between scholars.

Archaeological record provides evidence that Greek early Neolithic artefact have great analogies with the Middle-eastern and Anatolian ones. These analogies could suggest a model of regular advance of small communities, which founded new villages not far away from their previous settlements and left a homogeneous cultural landscape (Ammerman & Cavalli-Sforza 1984). Although, the parallels that can be established between Greece, the Middle-east and Anatolia come from different regions and even different periods. Moreover, the heterogeneous cultural landscape found between Middle-east and Greece suggests a model of rapid displacements of small groups over long distances, ultimately settling in favoured environments, far from their original homes (van Andel and Runnels 1987). The most striking formal artefactual analogies point alternately to the Levant, the Jordan valley, central Turkey and also Crete as the possible origin of the Neolithic migrants, making difficult to identify the geographical origin of the Greek farmers. The same uncertain picture it is drawn when the route that brought Neolithic farmers into mainland Greece is analysed. Archaeological records show a substantial homogeneity in the artefactual technology among the early Neolithic sites within Greece. Moreover, the slight differences in artefactual technologies found, among the different Greek areas, can be attributed to the impact of local Mesolithic groups, which were rapidly assimilated and have, obviously, promoted original technical and stylistic characteristics. Therefore, no clear evidence can be retrieved from the analysis of artefacts regarding the route followed by these long-distance migrants during the colonization of Greece (Perlès 2001).

To this regard three alternative theories have been proposed over the years (Figure 1.5.1a). The first two routes are both maritime. It is well established that navigation has been known in Greece since the Late Pleistocene, as indicated by the presence of Melian obsidian in the Final Pleistocene and Early Holocene levels from Franchthi (Renfrew and Aspinall 1990).The first route (❶ in Figure 1.5.1a) implies a first colonization of Crete, from Cyprus or the Mediterranean coasts of central Turkey, followed by a second step that brought Neolithic farmers from Crete to the North-eastern coast of Peloponnesus (Ozdogan 1997; Gronenborn 1999; Colledge et al., 2004).

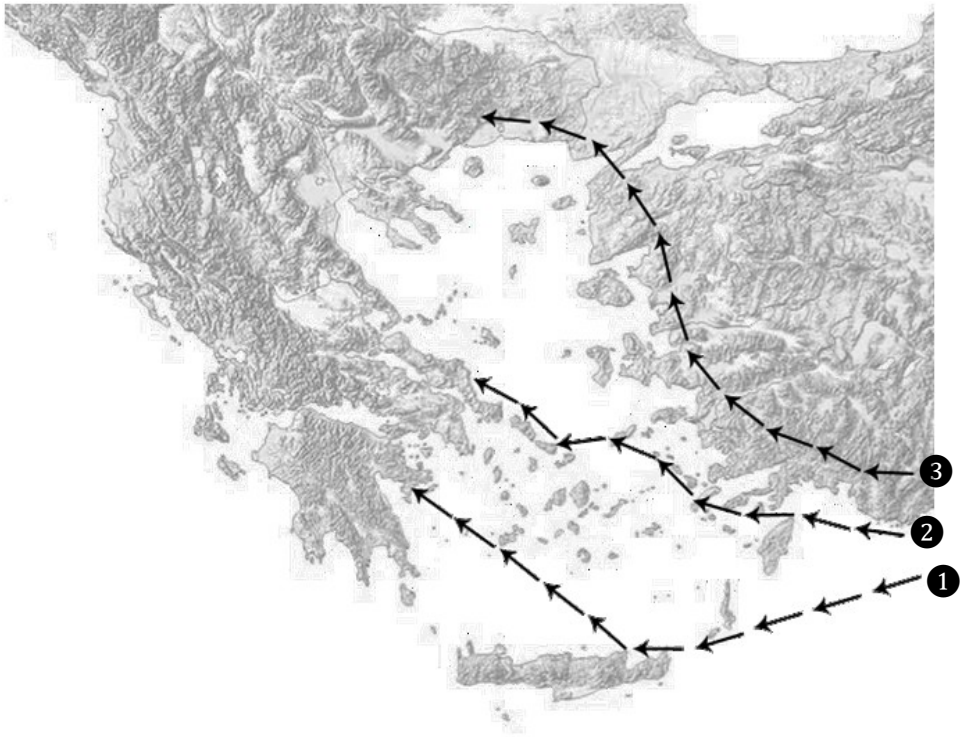



**Figure 1.5.1a**. Schematic visualization of the three possible routes used by Neolithic migrants. ❶= Cyprous/Crete/NE Peloponnesus route; ❷= Island hopping route; ❸=Inland route

The colonization of Crete it is well attested. Archaeological links have been previously drawn between the first settlers of Knossos and their central Anatolian predecessors/contemporaries, based on their common use of mud-brick technology and shared suites of domesticates, with particular reference made to bread wheat (*Triticum aestivum*). Moreover, some similarities in artefactual traditions are found between Cretan and North-eastern Peloponnesus early Neolithic (Ozdogan 1997).

This route it is also supported by a genetic survey carried out by King and colleagues in 2008. In this study, four Greek populations have been analysed, one from Crete and

three from mainland Greece close to well known early Neolithic sites in North-eastern Peloponnesus, Thessaly and Central Macedonia. Their conclusions are based on the frequency of haplogroup J2a-M410, previously associated with the Neolithic transition (Sengupta et al., 2006), found in high frequencies among the Cretan population and also in Central and Mediterranean Turkey. In mainland Greece, frequencies comparable to that obtained for Crete and Anatolia are found among the Southern sample (North-eastern Peloponnesus). The other two Northern samples, on the other hand, show low frequencies of the J2a-M410 lineage. This pattern, according to the authors, supports the Cyprus/Crete/NE Peloponnesus route. On the other hand, the presence in Crete of the J2a-M319, a sub-clade of J2a-M410 also found in Central and Mediterranean Turkey, at high frequencies and the virtual absence of this lineage in all the other mainland Greek samples could suggest a differential colonization of Crete and mainland Greece. Moreover, for North-eastern Peloponnesus there is evidence of local adoption of agriculture rather than colonization process (Dennel 1985), especially in the Franchthi cave site. In this site a mix of local and foreign traditions can be seen for lithic industries, faunal remains and also for ceramics.

The second route (❷ in Figure 1.5.1a) implies an island hopping route that brought Neolithic migrants into eastern shores of Central Greece (Euboea island and Thessaly) (Renfrew, 1987; Davis, 1992; Wijnen, 1993; Van Andel and Runnels, 1995; Perlès, 2001, Colledge et al., 2004). The high concentration of early Neolithic sites in these regions together with archaeological evidence, for instance the presence in Thessaly of monochrome painted wares which strongly reminisces the ones found in Turkey, support this migration route. On the other hand, wide-ranging trade networks have been considered to be an important incentive to the development and spread of agriculture (Runnels and van Andel 1988). Although, no artefacts are known to have been exchanged between Greece and the Near East during the early Holocene. Moreover, there is no definite early Neolithic settlement on the islands between Anatolia and Greece that supports this model of maritime colonization. Regardless of this, the absence of settlements on small islands may simply reflect the reluctance to settle in restricted environments with few potential resources. Genetic data provided by King et al., 2008 also rejects this scenario even if these data comes from just a sampling site in Thessaly and do not concern the Euboea island.

The third route hypothesis is the only one which considers a possible inland migration (Cavalli Sforza et al., 1984, Perlès 2001). This scenario (❸ in Figure 1.5.1a) is mainly supported by genetic evidences and more specifically by the "wave-of-advance" model proposed by Ammerman and Cavalli-Sforza in 1984. According these evidences, provided by the analysis of classical markers (Menozzi et al., 1978), and more recently, from ancient mtDNA (Haak et al., 2010) the route followed by these migrants was along the Turkish Thrace towards the Greek Thrace and Macedonia. Although, for this scenario no any archaeological evidence has been found, until now, that supports it. No Early Neolithic site has yet been found in Greek Thrace or eastern Macedonia, the logical passage for an inland penetration from the Anatolia to Greece. Even if this absence has often been attributed to the effects of deep alleviation (Ozdogan 1993), their density should not exceed the one observed in other Greek regions. Moreover, the few early Neolithic sites found in eastern and Western Macedonia, especially those north of the Aliakhmon river, display strong Balkan affinities. This suggests that they were probably settled from the north rather than from the south or the east (Perlès 2001).

# 2. AIMS OF THE STUDY

One of the turning points in the history of European populations was the adoption of agriculture technology that had as a consequence the transition from the subsistence economy based on hunting and gathering to that based on food producing (Ammerman and Cavalli-Sforza 1984). This process is known as the Neolithic transition. The first European country to be touched in the process of "Neolithization" was Greece. Neolithic farmers spread from Middle-east in Greece around 9Kya. (Price 2000) bringing along the "Neolithic package", from there the Neolithic package spread into Europe reaching Britain in less than 3000 years.

Archaeological studies reported a high concentration of early Neolithic sites in central mainland Greece (Euboea and Thessaly) and Northern Peloponnesus (Argolide and Corinthian district). Archaeological evidence along with archaeobotanical, linguistic and craniometric reports suggest two main scenarios of migration into Greece. Both imply maritime route, but they differ in an important point. The first suggests an island-hopping route from Turkey (Anatolia) or the Levantine coasts to mainland Greece (Renfrew, 1987; Davis, 1992; Wijnen, 1993; Van Andel and Runnels, 1995; Perlès, 2001, Colledge et al., 2004). On the other hand, the second theory (Ozdogan 1997; Gronenborn 1999; Colledge et al., 2004) suggest a Pre-pottery Neolithic dispersal to Cyprus, Crete and finally to North-Eastern Peloponnesus. A third scenario (Cavalli Sforza et al., 1984, Perlès 2001) which will not be further considered, involves an inland migration from Turkey through Northern Greece (Thrace and Macedonia). This hypothesis is not supported by archaeological evidences since very few early Neolithic sites have been found in these areas.

The aim of this project is to test, through the analysis of both Y-chromosome and mtDNA, of the two majorly accepted scenarios of colonization of Greece by Neolithic farmers. Such hypotheses have testable expectations that will be evaluated through appropriate statistical analyses.

The first scenario (IHr) implies a higher genetic affinity of the Euboea island population with Anatolian and Middle-eastern populations. In fact, an initial colonization of this area followed by a spread to the rest of Greece should have left more Anatolian/Middle-eastern genetic signs there and less in North-eastern Peloponnesus. Secondly, a higher contribution and internal diversity of the Neolithic haplogroups

should point towards a dispersal of such lineages from Euboea island indicating this region as the arrival point of the Anatolian/middle-eastern migration route. Moreover, the time since expansion estimates for both population and Neolithic haplogroups should be older in Euboea island than North-eastern Peloponnesus. The times since expansion should provide an indicating of an earlier adoption of the Neolithic way of life and an earlier arrival of the Neolithic lineages.

On the contrary, a higher genetic affinity with Anatolian/Middle-eastern populations of the North-eastern Peloponnesus population will point to the Cyprus/Crete/North-eastern Peloponnesus route (CCPr). Moreover, for this scenario to be considered the most probable, the contribution and internal diversity of the Neolithic haplogroups should be higher in North-eastern Peloponnesus rather than Euboea island. Thirdly, older times since expansion for both populations and Neolithic haplogroups should be estimated for North-eastern Peloponnesus rather than Euboea island.

# 3. MATERIALS AND METHODS

### 3.1 Population analysed

In the present study we analysed a total of 206 samples from two Greek populations: Euboea island (n=96) and Corinthian district (n=110). Samples were collected from apparently healthy and unrelated male donors after being adequately informed about the research's aims and after they agreed to participate by signing informed consent. Birth place of parents and grandparents were registered in order to classify each individual to a pre-determined geographic area.

**Euboea island sampling**

The sampling in Euboea island has been carried out thanks to the help and support of the blood donor department of the General hospital of Chalkida. The samples from Euboea island were collected in 8 different locations which were grouped in three distinct areas identified on the basis of geographical, archaeological and historical criteria, namely Northern Euboea (n=24), Central Euboea (n=37) and Southern-central Euboea (n=35) (see figure 3.1a for sampling locations and geographical areas and (see Table 3.1a for details). Samples were classified to one area if they had both grandfathers and grandmothers born in within the area.

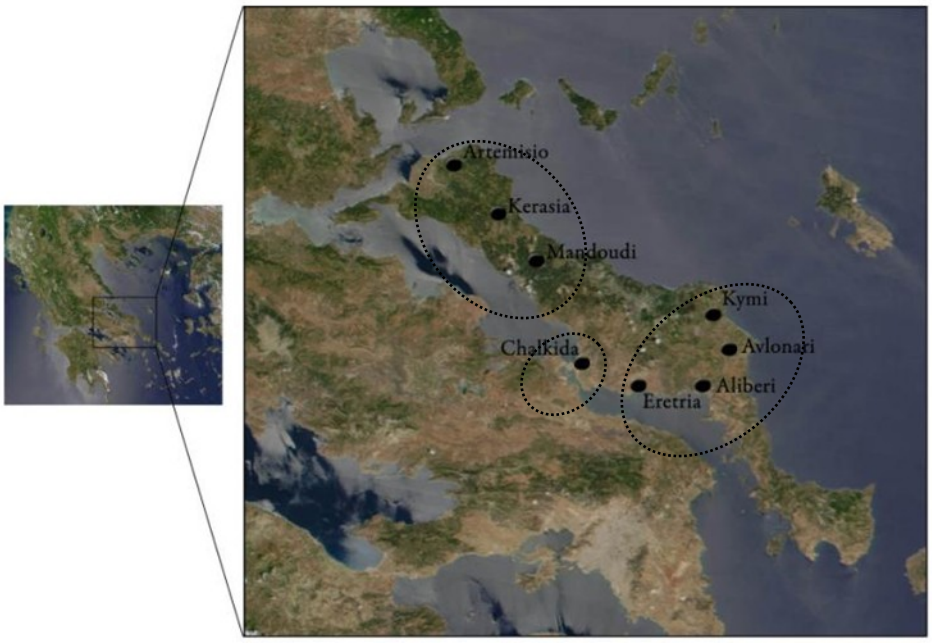



**Figure 3.1a**. Sampling locations (black dots) and geographical areas (dotted circles) in Euboea island.

**Table 3.1a**. Geographical position and number of samples collected for each sampling location.

| Area | Location | Longitude | Latitude | N |
|---|---|---|---|---|
| **NORTHERN EUBOEA** | Artemisio | 39°00'N | 23°13'E | |
| | Kerasia | 38°54'N | 23°19'E | |
| | Mandoudi | 38°47'N | 23°28'E | |
| | TOTAL | | | 24 |
| **CENTRAL EUBOEA** | Chalkida | 38°27'N | 23°35'E | 37 |
| | TOTAL | | | 37 |
| **SOUTHERN-CENTRAL EUBOEA** | Eretria | 38°23'N | 23°47'E | |
| | Aliveri | 38°24'N | 24°02'E | |
| | Avlonari | 38°30'N | 24°07'E | |
| | Kymi | 38°37'N | 24°06'E | |
| | TOTAL | | | 35 |

Southern Euboea was excluded from the sampling due to the presence, in this particular area, of a large community of Albanian speaking people, known as Arvanites, migrated in various regions of Greece during the late middle Ages (Hall 2000).

**Corinthian district sampling**

For the Corinthian district a total of 110 samples were collected in 4 different locations. Three distinct sampling areas were identified on the basis of geographical, archaeological and historical criteria, namely: North-eastern Korinthia (n=55), North-western Korinthia (n=34) and Southern Korinthia (n=21) within the region (see Figure 3.1b for sampling locations and geographical areas and table 3.1b for details). The same area classification rules used for Euboea were applied for Korinthia district. To avoid the collection of samples of Arvanite origin, since the distribution of these people in the Corinthian District is not localized as it is for Euboea island but they are dispersed all over the territory, a specific question has been added to the information questionnaire.

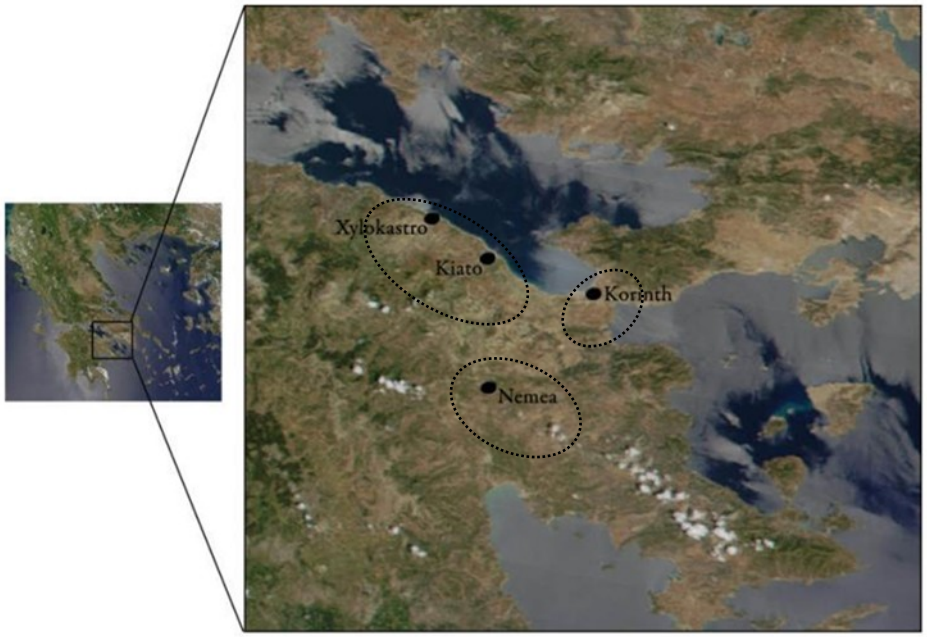



**Figure 3.1b**. Sampling locations (black dots) and geographical areas (dotted circles) in Korinthian district.

**Table 3.1b**. Geographical position and number of samples collected for each sampling location.

| Area | Location | Longitude | Latitude | N |
|---|---|---|---|---|
| **NORTH-EASTERN KORINTHIA** | Korinthia | 37°56'N | 22°55'E | 55 |
| | TOTAL | | | 55 |
| **NORTH-WESTERN KORINTHIA** | Kiato | 38°00'N | 22°44'E | |
| | Xylokastro | 38°04'N | 22°38'E | |
| | TOTAL | | | 34 |
| **SOUTHERN KORINTHIA** | Nemea | 37°49'N | 22°39'E | 21 |
| | TOTAL | | | 21 |

### 3.2 Laboratory analyses

In the present study both uniparental markers, mitochondrial DNA (MtDNA) and Y-chromosome (Y-chr) were analysed.

Regarding the MtDNA the hypervariable region 1 (HVR-I) was taken into consideration from position 16024 to position 16568 (according to Anderson et al., 1981). As for the Y-chr a total of 60 biallelic markers (SNPs) and 26 microsatellite loci (STRs) were genotyped. Genotyping of both MtDNA and Y-chr along with multiplex and probes design was carried out in the laboratory of the Department of Zoology of the University of Oxford by Dr. Francesca Brisighelli and supervised by Dr. Cristian Capelli. The sampling and genotyping was founded by the British Academy.

#### 3.2.1 DNA extraction

DNA samples were collected using buccal swabs; two samples for each donor were taken to ensure DNA availability. Dna was extracted using a modified "Salting out" method.

**Modified "Salting out" protocol for DNA extraction from buccal swabs**

Buccal swabs were collected using sterile brushes and stored inside 2 µl tubes at -20°C until extraction.

- Add 1 µl of ethanol (100%) and vortex the tubes to resuspend the cheek scraps cells
- Discard the brush and centrifuge the tubes with cheek scraps 5 min. at 14000 rpm. Discard supernatant
- Resuspend pellet in 200 µl **High TE**; vortex the tubes for 30 sec.
- Add 270 µl **Madisen Lysis Solution** and 10 µl **Proteinase K**; vortex the tubes for 30 sec.
- Incubate overnight at 37°C
- Ad 120 µl **NaCl** solution 3M; vortex the tubes and centrifuge 10 min. at 14,000 rpm.
- Remove supernatant to a new 1.5 µl tube and discard the pellet
- Add enough chilled ethanol (100%) to fill the tube
- Let the tubes overnight at -20°C
- Centrifuge 10 min. at 14,000 rpm. And discard supernatant

- Add enough ethanol (70%) to fill the tubes and centrifuge at 14,000 rpm. for 10 min. discard supernatant
- Dry the empty tubes 1 hour at 60°C and add 50 µl of water

Extraction products were visualized and quantified by electrophoresis on 1% agarose gel.

### 3.2.2 MtDNA genotyping

The first hypervariable region of the MtDNA of a total of 181 samples (96 for Euboea island and 86 for Corinthian District) was PCR-amplified using primers 15997L and 017H (Brandstatter et al 2004). Primer sequences are reported in table 3.2.2a.

**Table 3.2.2a**. Primers sequence used for HVR-I genotyping.

| | |
|---|---|
| 15997L | CACCATTAGCACCCAAAGCT |
| 017H | CCCGTGAGTGGTTAATAGGT |

PCR products were sequenced using the BigDye Terminator v3.1 Sequencing Standard kit following producer's instructions. All sequences were read from position 16024 to position 16568 and aligned using the software DNA Alignment 2.1 (Fluxus engineering). Haplogroup assignment was accomplished by the use of the Haplogrep haplogroup predictor (Brandstaetter et al., 2010; van Oven et al., 2009).

### 3.2.3 Y-Chromosome genotyping

- **SNPs genotyping**

All the 206 samples collected were genotyped for 60 SNPs following a hierarchical approach using a multiple single-base extension reaction approach with the SNaPshot® Multiplex Kit (Applied Biosystems) method. The first PCR step was performed with a QIAGEN Multiplex PCR kit which is used to obtain multilocus amplification products starting from a uniform primer concentration. The snapshot technique basically consists in a single base sequence. The probes for the sequencing are designed to stop just a base before the polymorphic site. The sequence reaction is blocked as soon as a single

marked nucleotide is integrated and the sequence is thus read to check which allele is present in the analysed individual.

Five multiplexes were designed for the hierarchical genotyping of the samples. The first multiplex consisted in 13 SNPs in order to define the basal European haplogroups (see Table 3.2.3a for details)

**Table 3.2.3a**. Mean features of the 13 SNPs for the basal multiplex.

| Marker | Accesion code | Haplogroup | Chromosome position | Base substitution | Reference |
|---|---|---|---|---|---|
| M269 | rs9786153 | R1b1b2 | 21148755 | T/C | Onofri et al., 2006 |
| M17 | rs3908 | R1a1a | 20192556 | C/G | Onofri et al., 2006 |
| M201 | rs2032636 | G | 13536923 | G/T | Onofri et al., 2006 |
| M170 | rs2032597 | I | 13357186 | A/C | Onofri et al., 2006 |
| M172 | rs2032604 | J2 | 13479028 | T/G | Onofri et al., 2006 |
| M35 | rs1179188 | E1b1b1 | 20201091 | G/C | Onofri et al., 2006 |
| M9 | rs3900 | K | 20189645 | C/G | Onofri et al., 2006 |
| M45 | rs2032631 | P | 20327175 | G/A | Onofri et al., 2006 |
| M173 | rs2032624 | R1 | 13535818 | A/C | Onofri et al., 2006 |
| M89 | rs2032652 | F | 20376701 | C/T | Onofri et al., 2006 |
| M267 | rs9341313 | J1 | 21151206 | T/G | Onofri et al., 2006 |
| M282 | rs13447371 | H2 | 21905719 | A/G | Onofri et al., 2006 |
| M304 | rs13447352 | J | 21159241 | A/C | Onofri et al., 2006 |

**Table 3.2.3b**. Mean features of the 46 SNPs for the 5 multiplexes.

| | | **HAPLOGROUP J2** | | | |
|---|---|---|---|---|---|
| **Marker** | **Accesion code** | **Haplogroup** | **Chr. position** | **Base substitution** | **Reference** |
| M410 | AC006040.3 | J2a | 2811678 | A/G | Sengupta et al., 2006 |
| M102 | rs2032608 | J2b | 20385500 | G/C | Onofri et al., 2006 |
| M67 | rs2032628 | J2a4b | 20338197 | A/T | Onofri et al., 2006 |
| M47 | AC009977 | J2a4a | 20210718 | G/A | Onofri et al., 2006 |
| M319 | rs13447373 | J2a4d | 13977179 | T/A | Shen et al., 2000 |
| M241 | rs8179022 | J2b2 | 13527853 | G/A | Cinnioglu et al., 2004 |
| M92 | rs2032648 | J2a4b1 | 20363411 | T/C | Underhill et al., 2001 |
| M280 | AC010889.3 | J2b2b | 20338150 | G/A | Semino et al., 2004 |
| | | **HAPLOGROUP I** | | | |
| **Marker** | **Accesion code** | **Haplogroup** | **Chr. position** | **Base substitution** | **Reference** |
| M253 | rs9341296 | I1 | 14031372 | C/T | Cinnioglu et al., 2004 |
| M438 | rs17307294 | I2 | 15148198 | A/G | Underhill et al., 2007 |
| M223 | AC003032 | I2b1 | 15208728 | C/T | Onofri et al., 2006 |
| M26 | rs2032629 | I2a1 | 20325209 | G/A | Onofri et al., 2006 |
| M423 | AC007034.4 | I2a2 | 17605485 | C/T | Underhill et al., 2007 |
| P37.2 | AC002992.1 | I2a | 13001692 | T/C | YCC, 2002 |
| | | **HAPLOGROUP E1b1b1a** | | | |
| **Marker** | **Accesion code** | **Haplogroup** | **Chr. position** | **Base substitution** | **Reference** |
| M107 | RS2032613 | E1b1b1b1 | 20391026 | A/G | Onofri et al., 2006 |
| M81 | RS2032640 | E1b1b1b | 20351960 | C/T | Onofri et al., 2006 |
| M165 | AC010889 | E1b1b1b2a | 20326047 | G/C | Onofri et al., 2006 |
| M123 | AC010889.3 | E1b1b1c | 20223974 | G/A | Underhill et al., 2007 |
| M281 | rs13447370 | E1b1b1d | 20223888 | G/A | Semino et al., 2002 |
| V6 | - | E1b1b1e | 6992007 | G/C | Cruciani et al., 2004 |
| P72 | AC010137.3 | E1b1b1f | 20070241 | G/A | Hammer et al., 2003 |
| | | **HAPLOGROUP E1b1b1** | | | |
| **Marker** | **Accesion code** | **Haplogroup** | **Chr.position** | **Base substitution** | **Reference** |
| V12 | - | E1b1b1a1 | 6883099 | A/G | Cruciani et al., 2006 |
| M78 | ac010889 | E1b1b1a | 20352691 | C/T | Onofri et al., 2006 |
| V13 | - | E1b1b1a2 | 6902263 | G/A | Cruciani et al., 2006 |
| M521 | - | E1b1b1a5 | 6882948 | C/T | Battaglia et al., 2008 |
| V19 | - | E1b1b1a3b | 20355588 | T/C | Cruciani et al., 2006 |
| V22 | - | E1b1b1a3 | 6919957 | T/C | Cruciani et al., 2006 |
| M224 | - | E1b1b1a1a | 20352687 | T/C | Underhill et al., 2001 |
| V32 | - | E1b1b1a1b | 6992821 | G/C | Cruciani et al., 2006 |
| V27 | - | E1b1b1a2a | 6956051 | A/T | Cruciani et al., 2006 |
| V65 | - | E1b1b1a4 | 15797066 | G/T | Cruciani et al., 2007 |
| M148 | ac010889 | E1b1b1a3a | 20355481 | A/G | Onofri et al., 2006 |
| | | **HAPLOGROUP R1b1b2** | | | |
| **Marker** | **Accesion code** | **Haplogroup** | **Chr. position** | **Base substitution** | **Reference** |
| M412 | rs9786140 | R1b1b2a1 | 8562236 | A/G | Busby et al, 2011 |
| S127 | rs9786076 | R1b1b2a1a | 16353412 | C/T | Busby et al, 2011 |
| S29 | rs17222279 | R1b1b2a1a1a | 15348893 | A/G | Busby et al, 2011 |
| S21 | rs16981293 | R1b1b2a1a1 | 8856078 | C/T | Busby et al, 2011 |
| S144 | rs7067305 | R1b1b2a1a2d3a | 12741292 | A/G | Busby et al, 2011 |
| S139 | rs2566671 | R1b1b2a1a2d3 | 5815550 | C/T | Busby et al, 2011 |
| M160 | - | R1b1b2a1a2d2 | 20348253 | C/A | Busby et al, 2011 |
| M126 | rs2032615 | R1b1b2a1a2d1 | 20389651-20389654 | Del 4bp | Busby et al, 2011 |
| S28 | rs1236440 | R1b1b2a1a2d | 13842543 | A/G | Busby et al, 2011 |
| S116 | rs34276300 | R1b1b2a1a2 | 20616699 | A/C | Busby et al, 2011 |
| M153 | - | R1b1b2a1a2b | 20165748 | T/A | Busby et al, 2011 |
| SRY 2627 | rs1800865 | R1b1b2a1a2c | 2718271 | C/T | Busby et al, 2011 |
| S145 | rs11799226 | R1b1b2a1a2f | 14163822 | C/G | Busby et al, 2011 |
| M222 | rs20321 | R1b1b2a1a2f2 | 13411808 | G/A | Busby et al, 2011 |

Following the hierarchical approach the haplogroups: J2, I, E1b1b1 (two multiplexes) and R1b1b2 were dissectioned in order to achieve a higher resolution and in depth phylogeny (see table 3.2.3b for details of all multiplexes). Data regarding the sub-typing of haplogroup R1b1b2 was retrieved from the paper of George Busby "Microsatellite choice and Y chromosome variation: the cautionary tale of "Neolithic" R-M269 lineage in Europe" which is under submission.

Two additional SNPs for the definition of the haplogroups G were genotyped by direct sequencing (see table 3.2.3c for details) using the BigDye Terminator v3.1 Sequencing Standard kit following producer's instructions.

**Table 3.2.3c**. Features and primer sequences of the 2 SNPs for haplogroup G-M201.

| Marker | Accesion code | Haplogroup | Primers | Base substitution | Reference |
|---|---|---|---|---|---|
| P15 | AC007876.2 | G2a | F- agagagttttctaacagggcg<br>R- tgggaatcacttttgcaact | C/T | Hammer et al., 2000 |
| M406 | - | G2a3a | F- ccccaaaaagctattctgtaa<br>R- gaagcactttagagcaggg | T/G | King et al., 2008 |

Haplogroup classification was based on the latest nomenclature (International Society of Genetic Genealogy (2011). Y-DNA Haplogroup Tree 2011, Version: 6.0, Date: 1-01-2011, http://www.isogg.org/tree/). A schematic phylogenetic tree based on the SNPs analysed can be visualized in Figure 3.2.3a.

- **STR genotyping**

A total of 26 STRs were genotyped for all samples. Seventeen Y chromosome STRs were typed using the AmpF*ℓ*STR® Yfiler® PCR Amplification Kit (Applied Biosystems) designed for loci: DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385 a/b, DYS393, DYS391, DYS439, DYS635, DYS392, Y GATA H4, DYS437, DYS438, DYS448.

For the other 9 STRs (DYS460, DYS388, DYS385a/b, YCA-II a/b, DYS461, DYS445, YGATA-A.10, DYS413 a/b) a multiplex reaction has been appositely designed (see Table 3.2.3d for details).

**Table 3.2.3d**. Features and primer sequences for STR multiplex.

| Marker | Repeat motif | Primer sequence | Reference |
|---|---|---|---|
| DYS460 | (ATG)n | F-agcaagcacaagaataccagag<br>R-tctatcctctgcctatcatttatta | Gusmao, Alves, 2005 |
| YCAIIa/b | (CA)n | F-tgtcaaaatttaacccacaatca<br>R-gcagtctttcaccataaggttagc | Butler et al, 2002 |
| DYS388 | (ATT)n | F-gaattcatgtgagttagccgtttagc<br>R-gaggcggagcttttagtgag | Butler et al, 2002 |
| YGATA-A.10 | (TCCA)2(TATC)n | F-cctgccatctctatttatcttgcatata<br>R-ataaatggagatagtgggtggatt | Gusmao, Alves, 2005 |
| DYS461 | (TAGA)11 (CAGA)n | F-aggcagaggatagatgatatggat<br>R-ttcaggtaaatctgtccagtagtga | Nist |
| DYS413a/b | (CA)n | F-aatgtgtgagccaattgtttagaa<br>R-gaaactaaaccaaacaggatactc | Malaspina,1997 |
| DYS385a/b | (GAAA) | 385.1-F-agcatgggtgacagagcta<br>385.2B-R-ccaattacatagtcctcctttc | Schneider et al, 1998 |
| DYS445 | (TTTA)n | F-gagctgagattatgccaccaaaa<br>R-agttaagagccccaccttcctg | Hanson et al 2006 |

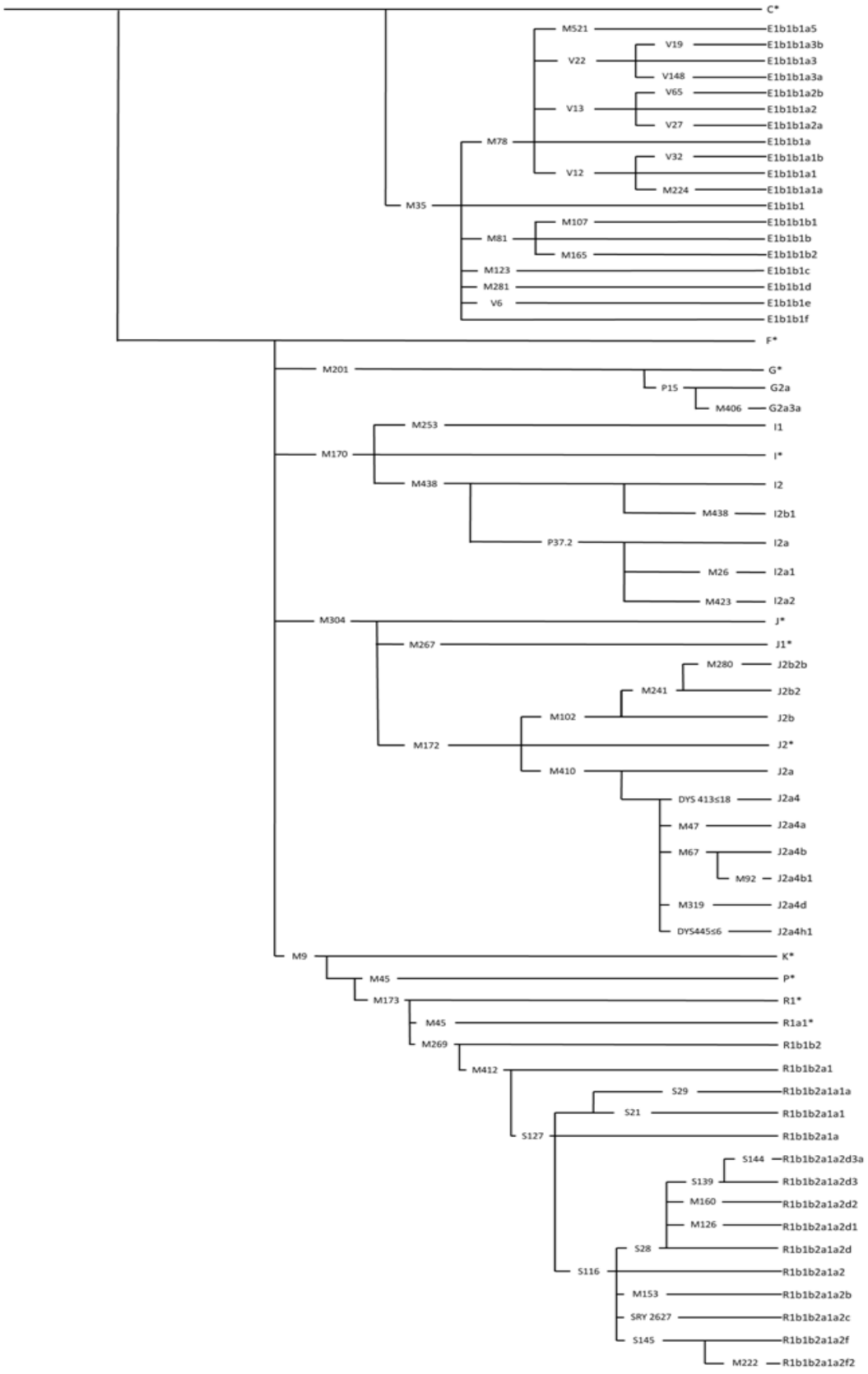



**Figure 3.2.3a**. Schematic phylogenetic tree based upon the 60 SNPs analysed in this thesis.

### 3.3 Statistical analyses

In the present section are described the statistical methods used for the data analyses.

#### 3.3.1 Intra-population variation

The first step in the analysis of the variation in the populations studied was the analysis of the intra-population variation.

The following parameters were studied:

- S, number of polymorphic sites: number of loci that present more than one allele per locus.

- k, number of different haplotypes: number of different haplotypes present in the population.

- Haplotype diversity: it is defined as the probability that two randomly chosen haplotypes are different in the sample and it is equivalent to the expected heterozygosity for diploid data. It is calculated as:

$$\hat{H} = \frac{n}{n-1}(1 - \sum_{i=1}^{k} p_i^2)$$

where $n$ is the number of gene copies in the sample, or of individuals in the case of an haploid

marker, $k$ is the number of haplotypes, and $p_i$ is the sample frequency of the $i$-th haplotype.

- MNPD, mean number of pairwise differences: it is the mean number of differences between all pairs of haplotypes in the sample. It is calculated as:

$$\hat{\pi} = \sum_{i=1}^{k} \sum_{j=1}^{k} p_i p_j \hat{d}_{ij}$$

where $d_{ij}$ is an estimate of the number of mutations having occurred since the divergence of haplotypes $i$ and $j$, $k$ is the number of haplotypes, and $p_i$ is the frequency of haplotype $i$.

- $\pi_n$, nucleotide diversity: it is the probability that two randomly chosen homologous nucleotides are different. It is equivalent to the gene diversity at the nucleotide level and it is calculated as the mean number of pairwise differences over the number of analysed loci:

$$\hat{\pi}_n = \frac{\sum_{i=1}^{k} \sum_{j<i} p_i p_j \hat{d}_{ij}}{L}$$

where $d_{ij}$ is an estimate of the number of mutations having occurred since the divergence of haplotypes $i$ and $j$, $k$ is the number of haplotypes, $p_i$ is the frequency of haplotype $i$ and L is the number of loci analysed.

- Mismatch Distribution: it is the distribution of the differences between all pairs of haplotypes in the sample and its mean value is π. Looking at the entire distribution and not only at its mean can allow a more precise description of the internal variation of the sample. In fact, the shape of the mismatch distribution has been observed to be indicative of population history, being influenced by population demographic state. A smooth, bell-shaped distribution indicates a period of rapid population growth, while a ragged, multimodal distribution indicates a population whose size has been constant over a long period (Rogers and Harpending, 1992). The raggedness is a parameter that allows to distinguish between these two types of distribution. This is the sum of the squared difference between neighbouring peaks and show higher values for multimodal distributions than for smooth ones.

- Analysis of the molecular variance (AMOVA): the analysis of variance of gene frequencies is a way to investigate population structure. The method implemented in Arlequin 3.5 (Excoffier et al., 2010) takes into account not only the frequencies but also the number of mutations between molecular haplotypes. In this way haplotype divergence information is integrated into the classical variance analysis.

The computation is meant to be hierarchically conducted on the basis of previous defined groups of individuals or populations. The variance is thus decomposed in covariant components due to intra-individual differences (c), inter-individual differences (b), and/or inter-population differences (a). For the haploid case, indeed, *i*-th haplotype frequency from the *j*-th population from the *k*-th group can be represented as a vector which is the sum of the covariance components and the unknown x component (frequency of *i*-th haplotype averaged over the whole).

$$\mathbf{x}_{ijk} = \mathbf{x} + \mathbf{a}_k + \mathbf{b}_{jk} + \mathbf{c}_{ijk} .$$

A Monte Carlo based test of permutation is applied to evaluate the significance of the observed values.

The intra-population diversity parameters were computed with the Arlequin 3.5 software (Excoffier et al., 2010)

### 3.3.2 Inter-population variation

#### 3.3.2.1 Genetic distances

The measures of genetic distance are statistics that allow to infer the evolutionary relations between populations or molecules. There are several methods to calculate genetic distances, with different assumptions and applications. In the present thesis have been used: the $F_{ST}$ method

The $F_{ST}$ method is a classic measure of genetic distance and can be applied to all kind of data. It varies between 0, for identical populations, and 1 for populations that share no alleles. It is based on allelic frequencies, it is calculated between pairs of populations and represents the excess of homozygotes in the subpopulations with respect to the metapopulation. It is calculated as follows:

$F_{ST} = Vp / p(1-p)$

where Vp is the variance of the frequency of the i allele in the metapopulation, and p is the mean frequency of the i allele between the populations.

The result of the analysis is a matrix of distances between populations that can be graphically represented.

### 3.3.2.2 Multidimensional scaling

The MultiDimensional Scaling (Kruskal, 1964) is a mathematical procedure that allows the representation of the objects under study in a Euclidean space, defined by a desired number of dimensions, so that the distances reproduced reflect the values observed in the best way possible. The method proceed through a series of iterations moving around objects in the space defined by the requested number of dimensions, and checking how well the distances between objects can be reproduced by the new configuration. The goal is to maximize the goodness-of-fit, which is represented by the stress value, defined as follows:

$$Phi = \Sigma (d_{ij} - \delta_{ij})_2$$

where $d_{ij}$ represents the observed distance between objects, while $\delta_{ij}$ is the reproduced distance. Higher the similarity between the two matrices, the observed and the reconstructed one, lower will be the value of stress. The iterations stop when this parameter cannot decrease further, reaching a threshold value.

The Multi dimensional scaling has been performed with the SPSS 16.0 software (release 16.0.1 for Windows, S.P.S.S. Inc.)

### 3.3.2.3 Phylogenetic reconstruction: Median-joining Network

The analysis of the intraspecific phylogenetic relationships has definitely improved in recent years. Traditional methods, such as maximum likelihood, maximum parsimony and minimum evolution have shown some limitations when analysing intraspecific data. This is due to specific phenomena that characterize the evolution of populations, such as sexual reproduction, recombination and small genetic distance between individuals, and to some specific characteristics of this kind of data, such as large sample size and homoplasy. For these reasons several network methods have been developed. These methods allow for reticulations which is the possibility for different trees to be represented in the same time (Posada and Crandall, 2001). The method used in the present work is the Median-Joining Network (Bandelt et al., 1999). This method is particularly appropriate when handling haploid data because it requires absence of recombination, and can be applied to both sequence data and STRs. The algorithm consists in two different phases:

it combines all minimum spanning trees (MSTs) in a unique network, generating a minimum spanning network; a MST is a tree that connects all the haplotypes without creating any reticulation and without introducing new nodes, constructed with a maximum parsimony principle. With a parsimonious criterion median vectors are added; these are consensus haplotypes of three close haplotypes, and are constructed considering three haplotypes at time. The median vectors can be interpreted biologically as possibly extant unsampled haplotypes or extinct ancestral haplotypes. Within the MJ calculations each position can be weighted according to its evolutionary rate.

In the present work the software Network 4.516 (Fluxus engineering) has been used to perform the Median-joining network

#### 3.3.2.4 Principal components analysis

Factor analysis is part of the general linear model (GLM) family of procedures. The approach consists in minimizing the number of factors or variables considered in a multifactorial analysis to obtain a graphical representation of the relationship among the variables. An intuitive example of this kind of reductionist approach is the linear regression that represents the best linear relation between pairs of observations of two distinct variables. When the number of the variables into account is more than two, the linear regression process starts building a straight that is the synthesis of variance for a first pair of variables, than it proceeds with the addition of the information belonging to a third variable. The first linear regression is thereby rotated to find out a new spatial orientation that maximizes the variance explained by a fictitious variable that is a synthesis of the three real variables and minimizes the variance in the surrounding space.

Principal Components Analysis (PCA) is based on this mechanism. The first component explains the maximum variance. Successive components explain progressively smaller portions of the variance and are all uncorrelated with each other. The PCA has been performed with SPSS 16.0 software (release 16.0.1 for Windows, S.P.S.S. Inc.).

### 3.3.3 Demographic inferences through coalescence Bayesian methods

- **Y Chromosome**

To estimate demographic parameters of both populations and haplogroups the BATWING software has been used (Wilson, Weale & Balding 2003). The ancestral effective population size, the time of the beginning of the populations' demographic expansion, and the growth rate were estimated. The software is based on a coalescent model and implements three different demographic models: constant population size, growing population size, and constant population size followed by demographic growth. The last one seems to be the most reasonable for populations that have undergone an agricultural revolution. Consequently, for all estimates the above-mentioned model was used.

Prior distributions were established to cover a range of expectations concordant with human population history (Wilson et al., 2003). For the effective population size, a gamma distribution (1; 0.0001) (Shi et al., 2010) was used, whereas for the alpha (growth rate) and beta (time at which expansion starts) priors, a gamma distribution of (2, 400) and (2, 1) (Balaresque et al., 2009), respectively, were used. For the analysis a subset of the 26 STRs, comprising 13 tetranucleotide STRs (DYS456, DYS389I/II, DYS390, DYS458, DYS19, DYS393, DYS391, DYS439, DYS635, GATA_H4, DYS437, DYS461), were chosen in order to obtain the most reliable evaluation of the posterior mutation rate distribution. A width mutation rate prior gamma distribution (1.47; 2130) was used for both STR loci sets as proposed by Shi et al., 2010. This distribution covers a range between $10^{-3}$ and $10^{-5}$ in accordance with the expected values of Y-chromosome STR mutation rate of both observed and effective estimates (see YHRD.ORG.3.0 for a summary of the main publications about Y chromosome STRs mutation rates, Zhivotovsky et al., 2004). Chain convergence was evaluated with two independent runs starting from different seeds. Number of sample was $1x10^{6}$ with treebetN=10 and Nbetsamp=20. The mode values of the posteriors distributions were calculated through R software package *modeest* (Poncet 2009, R Development Core Team 2009).

- **MtDNA**

The BEAST software, designed for molecular sequences, was used for demographic inferences on mtDNA (Drummond et al., 2005). This software can be applied to both

species phylogenies and population genetics. For population purposes, it is based on a coalescent model and it implements a Bayesian MCMC sampling algorithm as well as BATWING. However, the demographic models available in BEAST do not allow the same inferences achievable with BATWING. Under the model of constant population size followed by expansion, indeed, the times of population expansion are not provided. Considering this difference the Bayesian Skyline plot (BSP) was performed. This method estimates past population dynamics through time from a sample of molecular sequences without dependence on a pre-specified parametric model of demographic history. Uses a Markov chain Monte Carlo sampling procedure that efficiently samples a variant of the generalized skyline plot, given sequence data, and combines these plots to generate a posterior distribution of effective population size through time.

For population size and growth rate a uniform prior distribution of (10; 300000) and (0, 1) respectively was used. For the mutation rate a strict molecular clock option was used with a value of $4.125 \times 10^{-6}$ for generation, deduced from the recalibration of the human mtDNA molecular clock by Soares et al. (2009). The calibration of the mtDNA molecular clock was estimate through the time of divergence from other primate species and it represents a good evaluation of the evolutionary mutation rate. As for the BSP the same prior distribution for population size and mutation rates were used.

To evaluate convergence, as in the case Batwing, two independent run of $200 \times 10^6$ iterations were performed for both methods.

### 3.3.4 Time of the most recent common ancestor (T.M.R.C.A.) estimate

The age of microsatellite variation within a single lineage was estimated as the average squared difference in the number of repeats (ASD) between all current chromosomes and the founder haplotype, averaged over microsatellite loci defined as follows:

$$ASD = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{n}\sum_{j=1}^{n}(L_{ij} - L_i^0)^2\right)$$

The lack of haplotypes at high frequencies within many lineages made impossible to specify, with certainty, the founder. Therefore, to use an identical method for analysis of all lineages, the following approach was applied: for each locus, was computed the median value of repeat scores to form a median (rather than modal) haplotype, which

was then taken as a founder. This method gives an underestimate of microsatellite variation if the median haplotype deviates far from an actual founder. However, it is important to mention that the founding and median haplotypes coincide for a few hundred generations after the appearance of the lineage (Sengupta et al., 2006).

The resulting ASD value was than divided by $\mu=6.9\times10^{-4}$ per 25 years to estimate the T.M.R.C.A., with the SE computed over loci (Zhivotovsky et al. 2004; Zhivotovsky and Underhill 2005)

# 4. RESULTS

### 4.1 Y-chromosome results

For both Greek regions, Euboea island (Euboea) and Korinthia district (Korinthia) an AMOVA analysis, based on STR and SNP variation, was performed to assess whether the three areas sampled, for each region, could be pooled and used as a single meta-population. Results highlighted a low heterogeneity within each region (0.35% for STRs and 0.29% for SNPs) and a higher diversity between regions (0.5% for STRs and non significant for SNPs). Moreover a population differentiation test, performed for each region separately, showed no significant differentiation between areas of the same region. In the light of these results samples from the three areas of each region were pooled to form two meta-populations which were subsequently used for the analyses.

#### 4.1.1 Haplogroup frequencies and intra-population diversity

The analysis of the 60 biallelic markers allowed us to identify of 33 informative haplogroups. For simplicity haplogroups and sub-haplogroups names are reported with the first digits followed by the marker name (e.g. E1b1b1 defined by the M35 mutation becomes E1b-M35, the sub-branch of this haplogroup E1b1b1a2 defined by the mutation V13 becomes E1b-V13).

The main haplogroups observed in Europe (Semino et al., 2000) (E, I, J, R1a and R1b) also contribute to the gene pool of Euboea and Korinthia.

For both regions the haplogroup with the highest frequency is E1b-M35 (table 4.1.1a) and in particularly the sub-branch E1b-V13 which accounts for the majority of the E1b-M35 haplotypes (89.5% for Euboea and 83.4% for Korinthia) in accordance with this sub-branch's frequency values found in other Balkan populations (Battaglia et al., 2009). The haplogroup I is restricted in western Eurasia, showing high frequencies in the Balkans and in Scandinavia. The I1-M253 sub-branche is mostly found in the Northern Europeans while his sister clade I2-M438 is one of the most represented lineages in South-eastern Europe, reaching values of more than 35% in the Balkans. Within the Greek populations analysed the higher frequency of I2-M438 is observed in Korinthia (15.5%), while the frequency of this lineage in Euboea reaches slightly lower values (10.4%). It is worth to notice that the majority of I2-M438 lineages in Korinthia belong

to the I2a-M423 branch which accounts for the 64.7% of the total haplotypes. On the contrary Euboea does not show a predominant sub-branch for haplogroup I2-M438 (I2a-M423=40% and I2b-M223=50%).

The R1 haplogroup is common throughout Western Eurasia and account for more than 20-25% of the Y-chromosome pool in both Greek regions. With the exception of 3 R1-M173 individuals, all the remaining R1 lineages belong to R1a-M17 either R1b-M269. These two sub-clades, which show in Europe opposite frequency gradients with maximum incidences in eastern and western regions, respectively, still display high frequencies in the Greek regions. Haplogroup R1a-M17 shows a slightly higher frequency in Korinthia (17.3%) rather than Euboea (10.4%). Haplogroup R1b-M269 shows similar frequency for both Euboea and Korinthia (9.4% and 9.1% respectively) although the sub-structure of this haplogroup is different in the two regions. For Korinthia the majority of the haplotypes falls under the R1b-M269* branch (70%), whereas for Euboea the distribution resulted more homogeneous.

The J2-M172 is found in high frequencies in both populations. This lineage is also very frequent in the Middle-east and Anatolia. Both the sub-branches of haplogroup J2-M172 (J2a-M410 and J2b-M102) are found in the two Greek regions. The main branch is J2a-M410, a higher frequency of this haplogroup is observed for Euboea (15.6%) rather than Korinthia (10.9%). The inner structure of J2a-M410 reveals that the main branch for Euboea is J2a-M67, which accounts for the 40% of the total J2a-M410 haplotypes while the majority of the Korinthia haplotypes (75%) falls within the J2a-M67 and J2a-DYS445≤7 branches. The sister clade of J2a-M410, J2b-M102 shows a higher frequency in Euboea (9.4%), with the majority of haplotypes belonging to the J2b-M241 sub-branch (77.8%). On the other hand Korinthia shows a frequency of haplogroup J2b-M241 of 5.5%, with all haplotypes belonging to the J2b-M241 lineage. Finally it is worth notice that Euboea has a frequency of haplogroup G-M201 more than three times higher than Korinthia (9.5% and 2.7% respectively) with the main representative of this haplogroup being the G2a-P15 sub-branch (55.5%). Other minor lineages (C-M216, F-M89, I1-M253, J1-M267, K-M9 and P-M45) accounts for the remaining haplogroup diversity.

**Table 4.1.1a**. Y-chromosome haplogroup frequencies. Basal haplogroups in bold, sub-branches in italic.

| | EUBOEA(96) | | KORINTHIA(110) | |
|---|---|---|---|---|
| | N of haplotypes | Frequency | N of haplotypes | Frequency |
| **C-M216** | **1** | **0,010** | | |
| **E1b-M35** | **19** | **0,198** | **30** | **0,273** |
| *E1b-M78* | | | *1* | *0,009* |
| *E1b-V12* | | | *1* | *0,009* |
| *E1b-V13* | *17* | *0,177* | *25* | *0,227* |
| *E1b-V22* | *1* | *0,010* | *1* | *0,009* |
| *E1b-M123* | *1* | *0,010* | *2* | *0,018* |
| **F-M89** | **2** | **0,021** | **1** | **0,009** |
| **G-M201** | **9** | **0,094** | **3** | **0,027** |
| *G-M201** | *2* | *0,021* | | |
| *G2a-P15* | *5* | *0,052* | *1* | *0,009* |
| *G2a-M406* | *2* | *0,021* | *2* | *0,018* |
| **I1-M253** | **2** | **0,021** | **1** | **0,009** |
| **I2-M438** | **10** | **0,104** | **17** | **0,155** |
| *I2-M438** | *1* | *0,010* | *3* | *0,027* |
| *I2a-M26* | | | *1* | *0,009* |
| *I2a-M423* | *4* | *0,042* | *11* | *0,100* |
| *I2b-M223* | *5* | *0,052* | *2* | *0,018* |
| **J1-M267*** | **3** | **0,031** | **7** | **0,064** |
| **J2a-M410** | **15** | **0,156** | **12** | **0,109** |
| *J2a-DYS413≤18* | *2* | *0,021* | *2* | *0,018* |
| *J2a-M67* | *6* | *0,063* | *4* | *0,036* |
| *J2a-M92* | *3* | *0,031* | *2* | *0,018* |
| *J2a-DYS445≤7* | *4* | *0,042* | *4* | *0,036* |
| **J2b-M102** | **9** | **0,094** | **6** | **0,055** |
| *J2b-M102** | *1* | *0,010* | | |
| *J2b-M241* | *7* | *0,073* | *6* | *0,055* |
| *J2b-M280* | *1* | *0,010* | | |
| **K-M9** | **4** | **0,042** | **3** | **0,027** |
| **P-M45** | **1** | **0,010** | | |
| **R1-M173** | **2** | **0,021** | **1** | **0,009** |
| **R1a-M17** | **10** | **0,104** | **19** | **0,173** |
| **R1b-M269** | **9** | **0,094** | **10** | **0,091** |
| *R1b-M269** | *3* | *0,031* | *7* | *0,064* |
| *R1b-S139* | *3* | *0,031* | *1* | *0,009* |
| *R1b-S116* | *1* | *0,010* | | |
| *R1b-S21* | *2* | *0,021* | | |
| *R1b-S28* | | | *1* | *0,009* |
| *R1b-SRY2627* | | | *1* | *0,009* |

In a micro-regional scale the haplogroup structure looks quite similar for each area within the two regions (Figure 4.1.1a and Table 4.1.1b). It is worth noting that haplogroup J2a-M410, in the Euboea areas, shows a much higher frequency in the Central and Southern-central areas rather than the northern one (16.6%, 20% and 8.3% respectively). This latter value is in accordance with the frequency of this haplogroup in Thessaly, which is geographically very close to Northern Euboea (King et al., 2008). Moreover it is interesting to point out the high frequency of haplogroup R1a-M17 observed in the two Northern areas of Korinthia (20% and 24% for North-eastern and North-western Korinthia respectively) whereas in the Southern area this haplogroup is virtually absent.

Intra-population diversity parameters, calculated for both haplogroups and haplotypes, are reported in Table 4.1.1c. Haplogroup diversity resulted slightly higher in Euboea then in Korinthia for both the whole region and the areas within it. On the contrary, Korinthia exhibits a higher STR diversity (calculated as the mean variance over all 26 loci analysed). Haplotype diversity is extremely high and similar for all populations as expected with such a high number of STRs analysed.

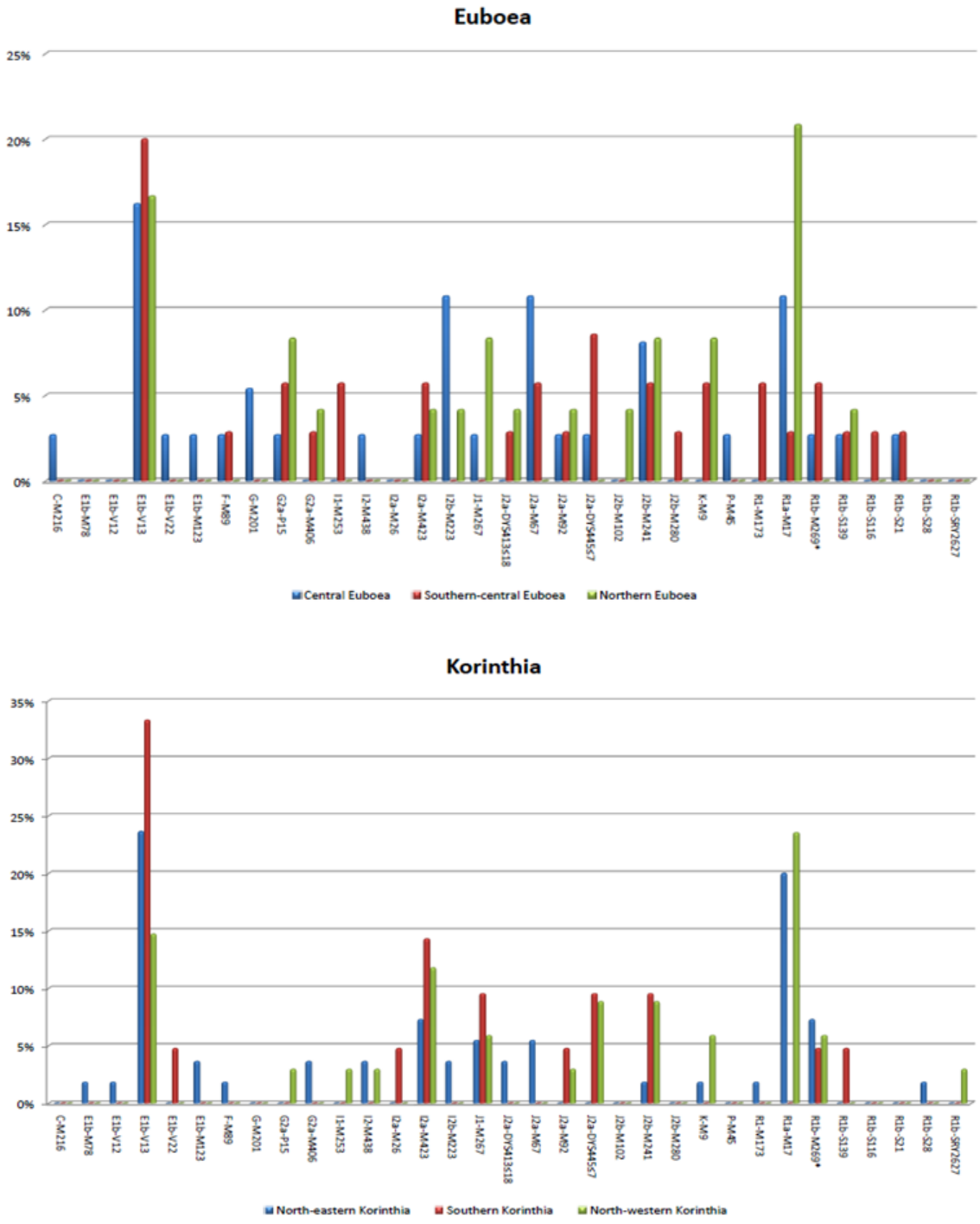



**Figure 4.1.1a**. Haplogroup structure and frequencies of the three areas within each region.

**Table 4.1.1b**. Haplogroup frequencies of the three areas within each region.

| | EUBOEA | | | KORINTHIA | | |
|---|---|---|---|---|---|---|
| | C-Euboea | SC-Euboea | N-Euboea | NE-Korinthia | S-Korinthia | NW-Korinthia |
| C-M216 | 0,03 | - | - | - | - | - |
| E1b-M78 | - | - | - | 0,02 | - | - |
| E1b-V12 | - | - | - | 0,02 | - | - |
| E1b-V13 | 0,16 | 0,20 | 0,17 | 0,24 | 0,33 | 0,15 |
| E1b-V22 | 0,03 | - | - | - | 0,05 | - |
| E1b-M123 | 0,03 | - | - | 0,04 | - | - |
| F-M89 | 0,03 | 0,03 | - | 0,02 | - | - |
| G-M201 | 0,05 | - | - | - | - | - |
| G2a-P15 | 0,03 | 0,06 | 0,08 | - | - | 0,03 |
| G2a-M406 | - | 0,03 | 0,04 | 0,04 | - | - |
| I1-M253 | - | 0,06 | - | - | - | 0,03 |
| I2-M438 | 0,03 | - | - | 0,04 | - | 0,03 |
| I2a-M26 | - | - | - | - | 0,05 | - |
| I2a-M423 | 0,03 | 0,06 | 0,04 | 0,07 | 0,14 | 0,12 |
| I2b-M223 | 0,11 | - | 0,04 | 0,04 | - | - |
| J1-M267 | 0,03 | - | 0,08 | 0,05 | 0,10 | 0,06 |
| J2a-DYS413≤18 | - | 0,03 | 0,04 | 0,04 | - | - |
| J2a-M67 | 0,11 | 0,06 | - | 0,05 | - | - |
| J2a-M92 | 0,03 | 0,03 | 0,04 | - | 0,05 | 0,03 |
| J2a-DYS445≤7 | 0,03 | 0,09 | - | - | 0,10 | 0,09 |
| J2b-M102 | - | - | 0,04 | - | - | - |
| J2b-M241 | 0,08 | 0,06 | 0,08 | 0,02 | 0,10 | 0,09 |
| J2b-M280 | - | 0,03 | - | - | - | - |
| K-M9 | - | 0,06 | 0,08 | 0,02 | - | 0,06 |
| P-M45 | 0,03 | - | - | - | - | - |
| R1-M173 | - | 0,06 | - | 0,02 | - | - |
| R1a-M17 | 0,11 | 0,03 | 0,21 | 0,20 | - | 0,24 |
| R1b-M269* | 0,03 | 0,06 | - | 0,07 | 0,05 | 0,06 |
| R1b-S139 | 0,03 | 0,03 | 0,04 | - | 0,05 | - |
| R1b-S116 | - | 0,03 | - | - | - | - |
| R1b-S21 | 0,03 | 0,03 | - | - | - | - |
| R1b-S28 | - | - | - | 0,02 | - | - |
| R1b-SRY2627 | - | - | - | - | - | 0,03 |

**Table 4.1.1c**. Intra-population diversity indices for Y-chromosome data based on 26 STRs. MNPD acronym stands for Mean Number of Pairwise Differences.

| | N | N haplogroups | Haplogroup diversity | N haplotypes | Haplotype diversity | MNPD | Mean variance |
|---|---|---|---|---|---|---|---|
| **EUBOEA** | **96** | **28** | **0,934 (+/-0,012)** | **95** | **0,9998 (+/-0,0015)** | **17,634 (+/-7,900)** | **1,427 (+/-0,221)** |
| C-Euboea | 37 | 20 | 0,944 (+/-0,019) | 37 | 1,000 (+/-0,006) | 17,490 (+/-7,949) | 1,325 (+/-0,197) |
| SC-Euboea | 35 | 19 | 0,946 (+/-0,023) | 35 | 1,000 (+/-0,007) | 17,876 (+/-8,129) | 1,553 (+/-0,278) |
| N-Euboea | 24 | 13 | 0,927 (+/-0,032) | 23 | 0,996 (+/-0,013) | 17,707 (+/-8,149) | 1,444 (+/-0,223) |
| **KORINTHIA** | **110** | **23** | **0,899 (+/-0,016)** | **107** | **0,9995 (+/-0,0013)** | **16,925 (+/-7,587)** | **1,558 (+/-0,237)** |
| NE-Korinthia | 55 | 18 | 0,895 (+/-0,025) | 54 | 0,999 (+/-0,004) | 16,514 (+/-7,467) | 1,618 (+/-0,291) |
| S-Korinthia | 21 | 10 | 0,871 (+/-0,057) | 21 | 1,000 (+/-0,015) | 17,038 (+/-7,895) | 1,538 (+/-0,295) |
| NW-Korinthia | 34 | 14 | 0,909 (+/-0,023) | 34 | 1,000 (+/-0,007) | 17,157 (+/-7,821) | 1,446 (+/-0,209) |

### 4.1.2 Demographic inferences

One of the main consequences expected from the adoption of agriculture was the rapid and strong demographic expansion, which should have left a detectable genetic footprint (Ammerman & Cavalli-Sforza 1984; Boyle & Renfrew 2000). In order to evaluate the expansion time in the two Greek regions a Bayesian method implemented in the Batwing software was used. The analysis was performed only on the pooled samples of each region in order to narrow as much as possible the confidence intervals of the estimates. A demographic model of constant population size followed by population growth was used.

For both regions the expansion times were compatible with the advent of Neolithic culture in Greece. For Euboea the time since expansion resulted slightly older (~8,8Kya) than for Korinthia (~7Kya) suggesting an earlier introduction of agricultural techniques in this region (see Table 4.1.2a for details on approximate modal, mean and median posterior values for the main parameters estimated).

**Table 4.1.2a**. Posterior estimates of demographic parameters values obtained. Abbreviations Na=effective ancestral population size**,** $T_0$ = time since population growth started, r= population growth rate and T.M.R.C.A.= Time of the most recent common ancestor**.**

| | Na | Na (95%c.i.) | $T_0$ | $T_0$ (95%c.i.) | r | r (95%c.i.) | T.M.R.C.A | T.M.R.C.A (95%c.i.) |
|---|---|---|---|---|---|---|---|---|
| **Euboea** | | | | | | | | |
| *mean* | 68114 | | 24340 | | 0,006 | | 101575 | |
| *median* | 59161 | 20306-166595 | 16014 | 1369-192968 | 0,005 | 0,002-0,013 | 78514 | 12380-551633 |
| *modal* | 47979 | | **8817** | | 0,004 | | 53720 | |
| **Korinthia** | | | | | | | | |
| *mean* | 83228 | | 18931 | | 0,005 | | 122291 | |
| *median* | 75257 | 25699-186034 | 11570 | 1017-168574 | 0,004 | 0,001-0,011 | 98685 | 16111-605447 |
| *modal* | 65235 | | **6997** | | 0,003 | | 73817 | |

Time since expansion was also estimated for all the haplogroups that have been so far associated with Neolithic transition in Europe (J2a-M410, R1b-M269 and G2a-P15) and the haplogroups prevalent in the Greek populations, in order to evaluate which can be associated with the migrations from the Middle-east which brought the "Neolithic package" to Greece.

For this analysis the same demographic model used for populations was adopted. The Time of the most recent common ancestors (T.M.R.C.A.) values reported was computed with the ASD method. The T.M.R.C.A. estimates obtained with the coalescent approach are very similar with those obtained with the ASD method (data not showed) corroborating the goodness of the BATWING runs performed and the prior distributions chosen for the analysis. The results obtained for all haplogroups that had 6 or more representatives in at least one region are reported in Table 4.1.2b.

The majority of the haplogroups (R1a-M17, G2a-P15, I2-M438, J1-M267 and J2b-M102) shows times since expansion which ranges from approximately 4,5Kya to 2,7Kya, compatible with Bronze Age and the development of the Helladic civilizations, more specifically with the spread of Mycenaean culture (Montjoy 1998). In fact the times since expansion for this haplogroups result older in Korinthia which is the centre of origin of the Mycenaean culture. Two haplogroups, namely R1b-M269 and E1b-V13, shows times since expansion ranging from around 7Kya to 6Kya, compatibly with late Neolithic.

**Table 4.1.2b**. Posterior estimates of demographic parameters values obtained. Abbreviations Na=effective ancestral population size**,** $T_0$ = time since population growth started, r= population growth rate and T.M.R.C.A.= Time of the most recent common ancestor**.**

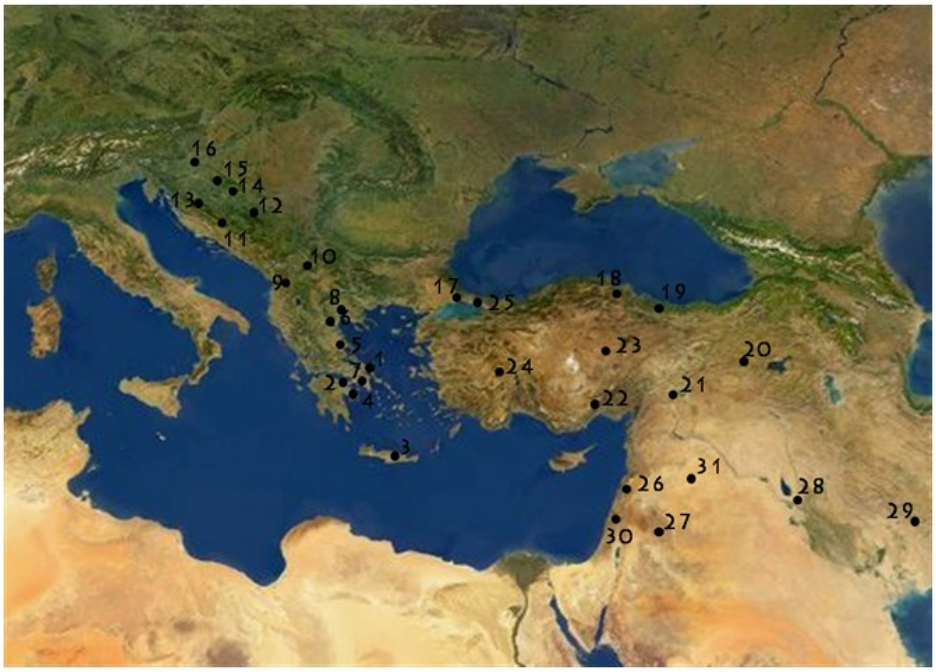
| | | Na | Na(95%c.i.) | $T_0$ | $T_0$(95%c.i.) | r | r(95%c.i.) | T.M.R.C.A |
|---|---|---|---|---|---|---|---|---|
| **E1b-V13** | | | | | | | | |
| | *mean* | 14837 | | 13206 | | 0,008 | | |
| Euboea | *median* | 13178 | 5439-33885 | 10326 | 343-115675 | 0,006 | 0,005-0,0023 | 11260(+/-3192) |
| | *modal* | 11139 | | **7125** | | 0,007 | | |
| | *mean* | 24380 | | 14267 | | 0,006 | | |
| Korinthia | *median* | 21405 | 7734-57705 | 10063 | 661-101746 | 0,006 | 0,002-0,015 | 13248(+/-4278) |
| | *modal* | 17439 | | **5973** | | 0,005 | | |
| **R1a-M17** | | | | | | | | |
| | *mean* | 28175 | | 12891 | | 0,005 | | |
| Euboea | *median* | 24406 | 8424-69934 | 8054 | 128-113031 | 0,004 | 0,001-0,013 | 12582(+/-3216) |
| | *modal* | 19732 | | **3036** | | 0,003 | | |
| | *mean* | 21496 | | 8710 | | 0,005 | | |
| Korinthia | *median* | 19380 | 7171-48076 | 5528 | 108-71416 | 0,004 | 0,001-0,014 | 9416(+/-2738) |
| | *modal* | 16568 | | **2277** | | 0,003 | | |
| **J2a-M410** | | | | | | | | |
| | *mean* | 47199 | | 27129 | | 0,005 | | |
| Euboea | *median* | 40499 | 14343-117738 | 18670 | 946-201831 | 0,004 | 0,001-0,012 | 30305(+/-8524) |
| | *modal* | 32323 | | **10134** | | 0,003 | | |
| | *mean* | 43122 | | 24451 | | 0,006 | | |
| Korinthia | *median* | 36837 | 12463-110260 | 16845 | 941-185434 | 0,006 | 0,001-0,015 | 23316(+/-4294) |
| | *modal* | 28882 | | **9384** | | 0,005 | | |
| **R1b-M269** | | | | | | | | |
| | *mean* | 41170 | | 24036 | | 0,005 | | |
| Euboea | *median* | 34628 | 11667-109930 | 16188 | 400-197995 | 0,004 | 0,001-0,012 | 19129(+/-3998) |
| | *modal* | 26857 | | **7262** | | 0,003 | | |
| | *mean* | 31757 | | 17636 | | 0,006 | | |
| Korinthia | *median* | 27269 | 9197-79752 | 11921 | 401-140921 | 5,000 | 0,001-0,015 | 15610(+/-3800) |
| | *modal* | 21658 | | **6314** | | 0,004 | | |
| **G2a-P15** | | | | | | | | |
| | *mean* | 44734 | | 19226 | | 0,005 | | |
| Euboea | *median* | 37283 | 12332-120679 | 11431 | 156-188319 | 0,004 | 0,001-0,013 | 19617(+/-5046) |
| | *modal* | 29193 | | **4162** | | 0,003 | | |
| Korinthia | | - | - | - | - | - | - | - |
| **I2-M438** | | | | | | | | |
| | *mean* | 34923 | | 14198 | | 0,005 | | |
| Euboea | *median* | 30382 | 10836-85279 | 9474 | 162-115328 | 0,004 | 0,001-0,013 | 19258(+/-4759) |
| | *modal* | 24746 | | **3736** | | 0,003 | | |
| | *mean* | 36341 | | 13989 | | 0,005 | | |
| Korinthia | *median* | 32452 | 11643-83345 | 9190 | 260-111381 | 0,004 | 0,001-0,013 | 14392(+/-3764) |
| | *modal* | 27071 | | **4605** | | 0,003 | | |
| **J1-M267** | | | | | | | | |
| Euboea | | - | - | - | - | - | - | - |
| | *mean* | 29449 | | 13839 | | 0,005 | | |
| Korinthia | *median* | 24065 | 7834-83025 | 8211 | 119-137353 | 0,004 | 0,001-0,014 | 10589(+/-3276) |
| | *modal* | 18746 | | **2956** | | 0,003 | | |
| **J2b-M102** | | | | | | | | |
| | *mean* | 24730 | | 12354 | | 0,005 | | |
| Euboea | *median* | 21149 | 7138-63478 | 7933 | 94-107621 | 0,005 | 0,001-0,014 | 9835(+/-2503) |
| | *modal* | 16890 | | **3023** | | 0,003 | | |
| | *mean* | 13852 | | 5807 | | 0,005 | | |
| Korinthia | *median* | 11157 | 3510-40060 | 3289 | 39-61721 | 0,004 | 0,001-0,014 | 4731(+/-1632) |
| | *modal* | 8437 | | **4013** | | 0,003 | | |

The only haplogroup that shows times since expansion compatible with early Neolithic is J2a-M410 in concordance with results reported in King et al., 2008.

Results obtained for haplogroup J2a-M410 points to an earlier expansion in Euboea (around 10Kya.) rather than Korinthia (around 9,4Kya). This observation seems to suggest that in Euboea the presence of migrant populations from the Middle-east and the consequent adoption of agriculture might have been achieved earlier than Korinthia.

### 4.1.3 Inter-population diversity

In order to define the genetic relationship of the two Greek regions analysed in this thesis with other Mediterranean (Greek and Balkan), Anatolian and Middle-eastern populations, an ad-hoc database of 29 different population has been constructed (see Figure 4.1.3a for details)




| | N |
|---|---|
| **1 Euboea** | 94 |
| **2 Korinthia** | 85 |
| **3 Crete** | 193 |
| **4 Lerna/Franchthi** | 57 |
| **5 Seskio/Dimini** | 57 |
| **6 Nea Nikomedia** | 57 |
| **7 Greeks** | 92 |
| **8 Mac-Greeks** | 57 |
| **9 Albanians** | 55 |
| **10 FYROM-Alb** | 64 |
| **11 Bosnia-Croats** | 90 |
| **12 Bosniacs** | 84 |
| **13 Croats** | 89 |
| **14 Bosnia-Serbs** | 81 |
| **15 Osijek-Croats** | 29 |
| **16 Slovenians** | 75 |
| **17 T1- Marmara** | 52 |
| **18 T2- W. Pontic** | 29 |
| **19 T3- E. Pontic** | 83 |
| **20 T4- E. Anatolia** | 82 |
| **21 T5- S/E. Anatolia** | 43 |
| **22 T6-Med. Anatolia** | 33 |
| **23 T7- C. Anatolia** | 90 |
| **24 T8- Aeg. Anatolia** | 30 |
| **25 T9- Istambul** | 81 |
| **26 Lebanon** | 916 |
| **27 Jordanian** | 146 |
| **28 Iraqy** | 203 |
| **29 Iranian** | 150 |
| **30 Palestinese** | 368 |
| **31 Syrians** | 202 |

**Figure 4.1.3a**. List of geographic location of population used for comparisons. Populations 1 and 2 are from present study. Populations 3-6 (King et al., 2008); 7-9 (Semino et al., 2004); 10, 15 and 16 (Battaglia et al., 2009); 11-14 (Marjanovic et al., 2005); 17-25 (Cinnioglou et al., 2004); 26,30-31 (Zalloua et al., 2008); 27 (Flores et al., 2005); 28 (Al-Zahery et al., 2003), (Sanchez et al., 2005); 29 (Regueiro et al., 2006).

A principal component analysis based on haplogroup frequencies has been carried out to explore the genetic affinities of the two Greek populations analysed with their neighbours (Figure 4.1.3b). Haplogroup resolution has been lowered in order to include a large number of populations.

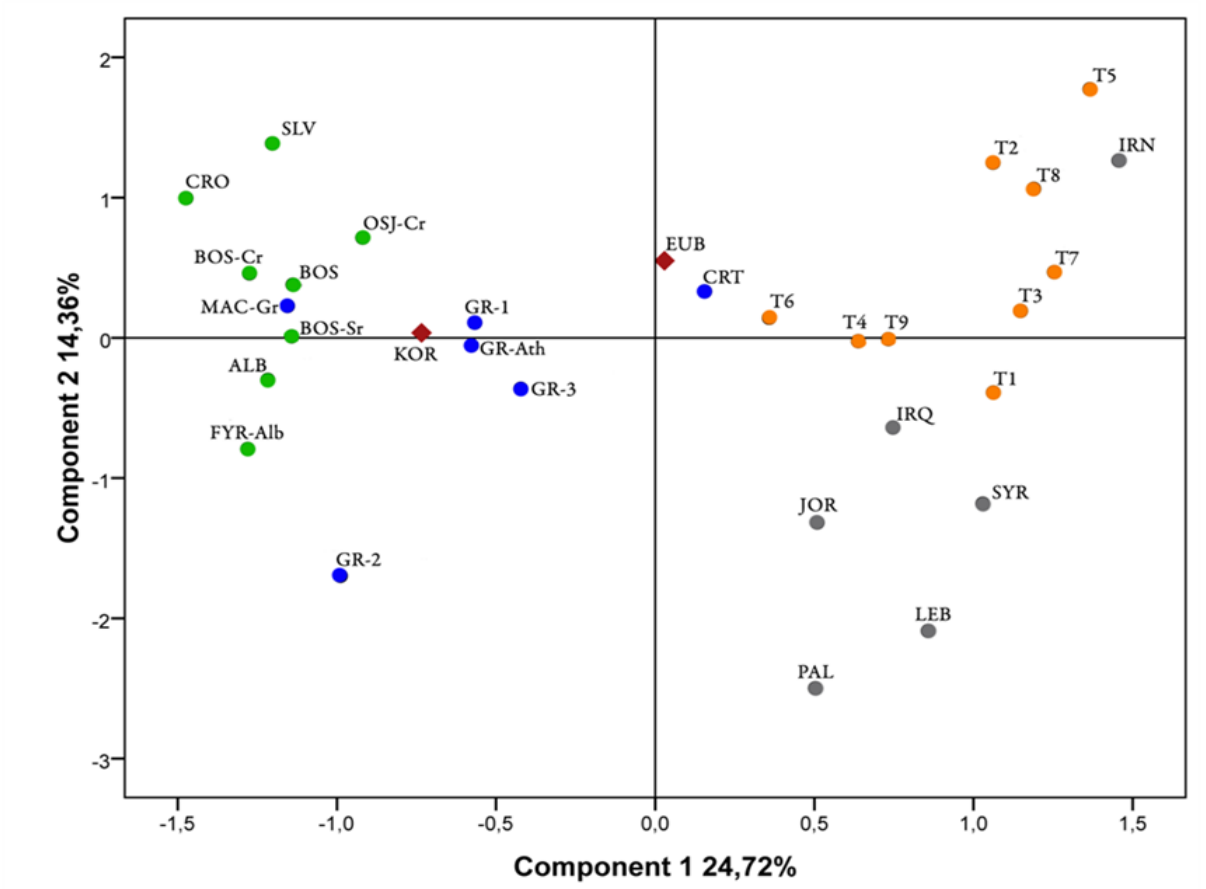



**Figure 4.1.3b**. Principal component analysis based on haplogroup frequencies. Abbreviations: SLV=Slovenians; CRO= Croats; OSJ-Cr=Osjiek Croats; BOS-Cr=Bosnia Croats; BOS=Bosnians; BOS-Sr= Bosnia Serbs; ALB= Albanians; FYR-Alb= FYROM Albanians; MAC-Gr= Macedonian Greeks (Thessalonica); KOR= Korinthia; GR-1=Nea Nikomedia; GR-Ath=Greeks from Athens; GR-2=Sesklo/Dimini; GR-3=Lerna/Franchthi cave; EUB=Euboea; CRT=Crete; T1-T9= Turks (see Figure 5.1.3a for details); IRN=Iranians; IRQ=Iraqi; SYR=Syrians; JOR=Jordanians; LEB=Lebanon; PAL=Palestinians

The total variance represented by the plot amounts to 39.08%. The first component accounts for the 24.74% of the total variance and the higher contribution to this component is given by haplogroup I2-M438 (~55%) followed by haplogroup J2a-M410 (~14%) and J1-M267 (~12%). This component separates the Middle-eastern and Anatolian populations (gray and orange circles respectively) from the Greek and Balkan ones (blue and green circles respectively) with the exception of two populations, Crete and Euboea which clusters with Anatolian/Middle-eastern group. Vector analysis (data not shown) demonstrates that the Greek/Balkan cluster is more associated with haplogroups I2-M438, J2b-M102 and R1a-M17; on the other hand, the Anatolia/Middle-

east cluster is more associated with haplogroups J2a-M410, J1-M267, G-M201 and R1b-M269. For the second principal component, which accounts for the 14,36% of the total variance, the higher contribution is given by haplogroup E1b-M35 which accounts for the 52% of the loading score. Moreover, the plot shows a closer relationship of Crete and Euboea with the Mediterranean Turkey (T6) and, in a lesser degree, with the Turkish sample from Istanbul (T9) and the Eastern Turkish sample (T4). The other Greek population analysed in this thesis, Korinthia, clusters with the other Greek and Balkan populations.

The multi dimensional scaling (MDS) plot based on the Fst genetic distance matrix calculated over the 7 STR loci (DYS19, DYS388, DYS390, DYS391, DYS392, DYS393 and DYS439) is presented in Figure 4.1.3c. Unfortunately, the analysis has been performed with a subset of (13) populations included in the database due to the lack of a complete STR profile in some populations of the database.

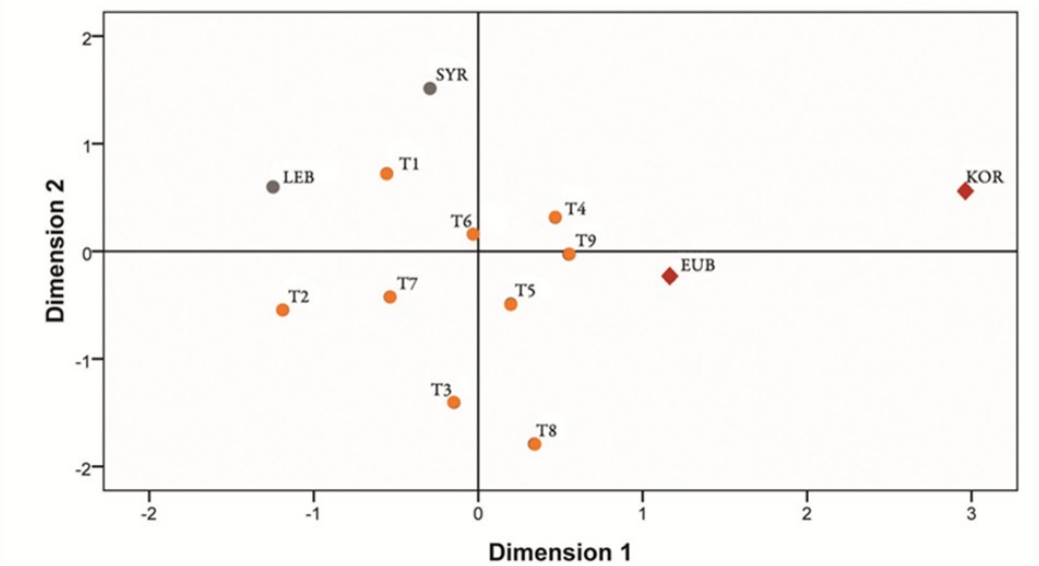



**Figure 4.1.3c**. Multidimensional scaling plot of the Fst genetic distances. The stress value (0.154) is acceptable according to Sturrocks and Rocha (2000). Statistically non significant values were converted to 0.

The plot points to a close genetic relationship of the Euboea sample with some Anatolian populations, whereas the Korinthia sample behaves as an outlier within this restricted database (see Table 4.1.3a for genetic distances values).

**Table 4.1.3a**. Matrix of the genetic distances between populations. Abbreviations for populations names are as in Figure 5.1.3b. Statistically non significant values have been converted to 0.

| | EUB | KOR | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9 | LEB | SYR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EUB | 0 | | | | | | | | | | | | |
| KOR | 0,014 | 0 | | | | | | | | | | | |
| T1 | 0,014 | 0,042 | 0 | | | | | | | | | | |
| T2 | 0,026 | 0,072 | 0 | 0 | | | | | | | | | |
| T3 | 0,013 | 0,052 | 0,022 | 0 | 0 | | | | | | | | |
| T4 | 0 | 0,022 | 0 | 0 | 0,012 | 0 | | | | | | | |
| T5 | 0 | 0,028 | 0 | 0 | 0 | 0 | 0 | | | | | | |
| T6 | 0 | 0,031 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | |
| T7 | 0,016 | 0,048 | 0 | 0 | 0,011 | 0 | 0 | 0 | 0 | | | | |
| T8 | 0 | 0,049 | 0,027 | 0 | 0 | 0,018 | 0 | 0,021 | 0 | 0 | | | |
| T9 | 0 | 0,017 | 0 | 0 | 0,013 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| LEB | 0,025 | 0,056 | 0 | 0 | 0,019 | 0,008 | 0,014 | 0 | 0 | 0,038 | 0,014 | 0 | |
| SYR | 0,024 | 0,054 | 0,008 | 0,023 | 0,028 | 0,015 | 0,011 | 0 | 0,012 | 0,039 | 0,009 | 0,009 | 0 |

### 4.1.4 Inter-lineage diversity: haplogroup J2a-M410

The haplogroup J2a-M410 has been often associated to the Neolithic Transition in Europe (Sengupta et al., 2006; King et al., 2008; Battaglia et al., 2009). Moreover this haplogroup was the only one that showed times since expansion compatible with the advent of the Neolithic in Greece for the two regions taken into consideration in this thesis (see paragraph 4.1.2). In the light of this evidence an in-depth investigation has been carried out for this lineage.

The two regions analysed the dissection of haplogroup J2a-M410 revealed only four informative sub-clades. The intra-lineage structure and internal diversity for Euboea, Korinthia and the 3 areas of both regions are showed in Figure 4.1.4a and Table 4.1.4a respectively.

Haplogroup diversity is similar for both regions. The most representative sub-clade for Euboea resulted to be J2a-M67*, which account for the 40% of the total J2a-M410 chromosomes. For Korinthia two sub-clades (J2a-M67* and J2a-DYS445≤7) account equally for the two thirds of the total J2a-M410 chromosomes. At a microregional level the only area within the two regions which shows all four haplogroups is Central Euboea which also shows the highest haplogroup diversity among the 6 areas. The Korinthia areas highlights the lower values of haplogroup diversity since all three of them exhibits only two J2a-M410 sub-clades (J2a-M92 and J2a-DYS445≤7 for Southern and North-western Korinthia; J2a-M67* and J2a- DYS413≤18 for North-eastern Korinthia).

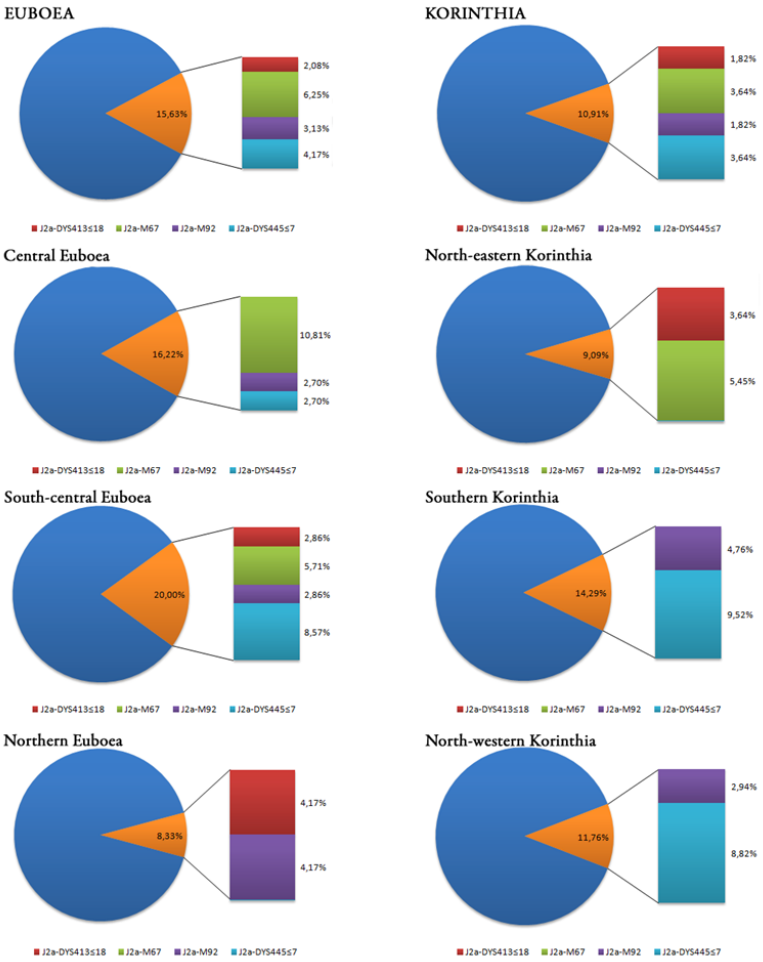



**Figure 4.4.1a**. Sub-clade structure of haplogroup J2a-M410 (orange slices) in the two Greek regions and in the 3 areas of each region.

**Table 4.1.4a**. Haplogroup J2a-M410 intra-lineage diversity. All 26 microsatellites were used for the estimates. Mean variance estimated only when 10 or more haplotypes were present. MNPD acronym stands for Mean Number of Pairwise Differences.

| | N | Haplogroup | N | Haplogroup | N | Haplotype | MNPD | Mean |
|---|---|---|---|---|---|---|---|---|
| **EUBOEA** | **15** | **0.156** | **4** | **0,762 (+/-0,066)** | **15** | **1,000 (+/-0,0024)** | **13,162 (+/-6,279)** | **0,933 (+/-0,228)** |
| C-Euboea | 6 | 0.162 | 3 | 0,600 (+/-0,215) | 6 | 1,000 (+/-0,096) | 13,600 (+/-7,141) | - |
| SC-Euboea | 7 | 0.200 | 4 | 0,809 (+/-0,129) | 7 | 1,000 (+/-0,076) | 12,428 (+/-6,339) | - |
| N-Euboea | 2 | 0.087 | 2 | 1,000 (+/-0,500) | 2 | 0,996 (+/-0,500) | 13,000 (+/-9,539) | - |
| **KORINTHIA** | **12** | **0.106** | **4** | **0,772 (+/-0,083)** | **12** | **1,000 (+/-0,0034)** | **14,182 (+/-76,847)** | **0,867 (+/-0,160)** |
| NE-Korinthia | 5 | 0.091 | 2 | 0,600 (+/-0,175) | 5 | 1,000 (+/-0,126) | 14,300 (+/-7,752) | - |
| S-Korinthia | 3 | 0.143 | 2 | 0,667 (+/-0,314) | 3 | 1,000 (+/-0,272) | 14,333 (+/-8,917) | - |
| NW-Korinthia | 4 | 0.177 | 2 | 0,500 (+/-0,265) | 4 | 1,000 (+/-0,177) | 12,833 (+/-7,361) | - |

Mean STR variance was calculated for regions only since the number of haplotypes in the areas were too small for a reliable estimate. The highest internal diversity has been detected for Euboea (0,933), while Korinthia shows a slightly lower internal diversity (0.867).

In order to investigate the phylogenetic relationship within the haplogroup J2a-M410 of the two Greek regions and Middle-eastern, Anatolian, Cretan and other Greek populations a network based on the 7 STRs (DYS19, DYS388, DYS390, DYS391, DYS392, DYS393 and DYS439) was built (Figure 4.4.1b). Each STR was given a specific weight according to its variance within the haplogroup: the weight of the ith STR was calculated as 10Vm/Vi, where Vm is the mean variance of all STRs and Vi is the variance of the ith STR.

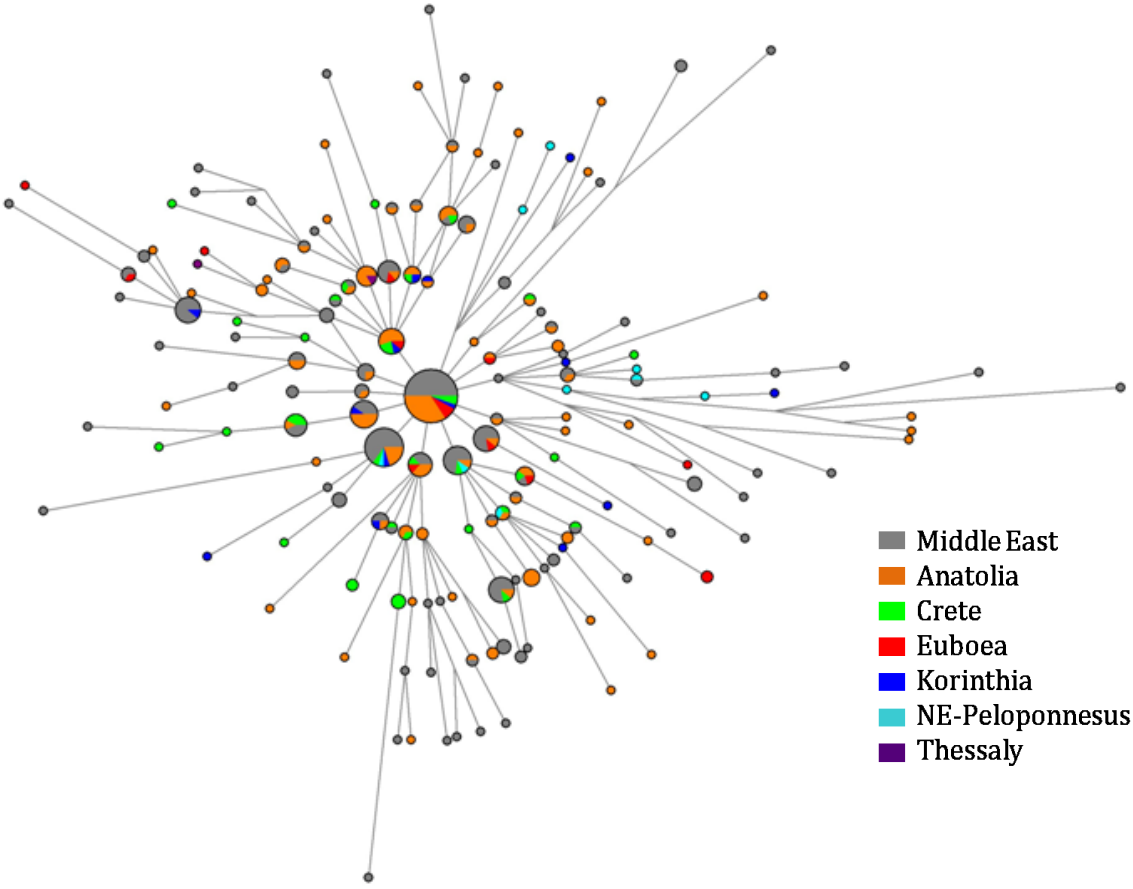



**Figure 4.4.1b**. Phylogenetic network of J2a-M410 haplogroup with individuals assigned to populations.

The network shows a simple star-like structure indicative of expansion from one source. The modal haplotype, as expected, is mainly composed by Middle-eastern followed by Anatolian haplotypes. A small percentage, around 30%, of the total modal haplotypes are represented by Cretan, Euboean and Korinthian ones with the Euboean

haplotypes being the majority (3 haplotypes; Korinthia 1 haplotype; Crete 2 haplotypes). Focusing on the one step mutation neighbours from the modal haplotype the contribution of Euboean chromosomes is higher than Korinthian ones for both number and relative frequency. In general, Euboea shares a higher number of haplotypes with Middle-east and Anatolia (8 and 9 out of 15 respectively) rather than Korinthia (5 and 6 out of 12 respectively).

The frequency distribution of the J2a-M410 lineage in a wide Eurasian context (Figure 4.1.4a)

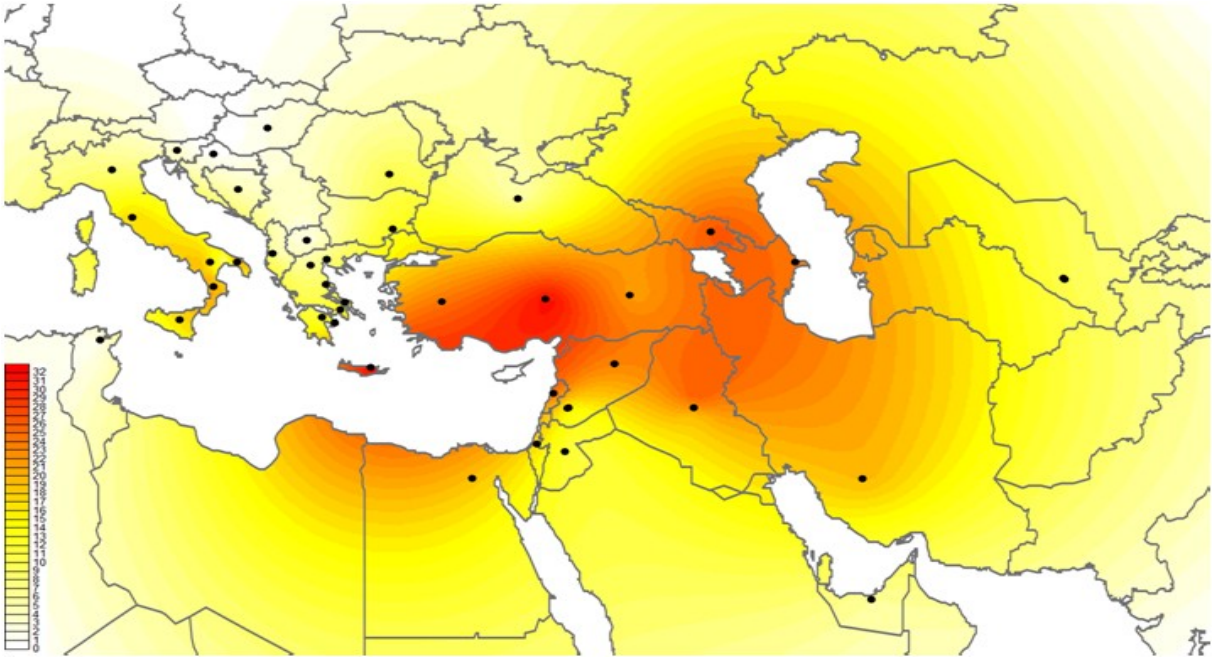


**Figure 4.1.4a**. Contour map of the distribution of haplogroup J2a-M410 in Eastern Eurasia. To the already described populations several others were added (Italian, North African, Balkanic and Caucasian)( Semino et al., 2004; Di Giacomo et al., 2004; Luis et al., 2004).

The distribution pattern highlighted by the contour map for haplogroup J2a-M410 seems to mirror, in some degree, the demographic spread of Neolithic farmers from the Middle-east towards Europe westwards and the towards the Indian sub-continent eastwards (Sengupta et al., 2006).

Considering only the Greek territory, there are two centres with the highest frequency distribution of haplogroup J2a-M410 (Figure 4.1.4b): Euboea island, mainly the central areas, and Crete. On the other hand the distribution of microsatellite diversity, exhibiting higher values in Euboea, suggests a possible dispersal of this haplogroup from this latter region and more precisely from the central areas.

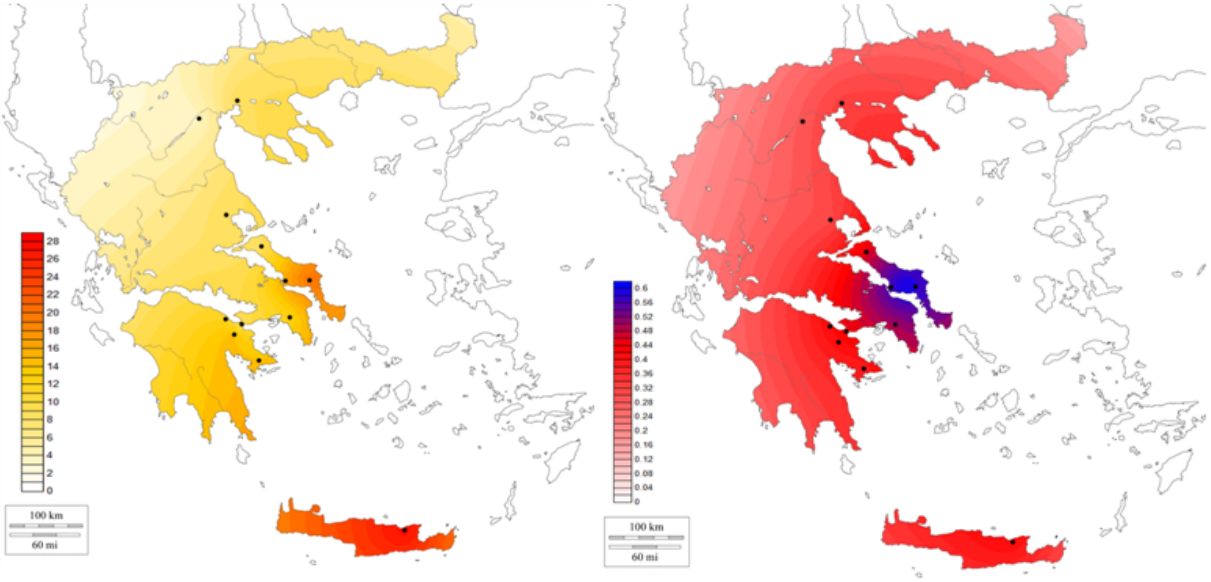


**Figure 4.1.4b**. Contour map of the frequency and variance distribution in Greece of haplogroup J2a-M410. For the mean STR variance only populations with 5 or more J2a-M410 haplotypes were considered. For Crete the J2a-M319 haplotypes were excluded due to their outlier microsatellite pattern (Malaspina et al., 2001; Martinez et al., 2007).

### 4.2 Mitochondrial DNA

From the total 206 samples, 181 (95 for Euboea and 86 for Korinthia) were genotyped for the first hypervariable region of mitochondrial DNA (MtDNA). An AMOVA analysis has been carried out in order to assess whether the three areas samples could be pooled together. The amount of variance between regions was 0.91% while the variance between areas within regions was slightly higher (0.93%). Despite this, being the differences between areas only slightly higher than the differences between regions, the pooled sample has been also used for analyses.

#### 4.2.1 Intra-population diversity

Being the information for diagnostic sites located outside the HVR-I region not available, the Haplogrep software was used for haplogroup assignment. This software, developed by the University of Innsbruck, provides, the best estimate for haplogroup classification based on the latest Phylotree release, through the use of Bayesian statistics. All samples exhibited more than 70% score of correct haplogroup classification. The haplogroups that showed the highest uncertainties, the CRS (Cambridge Reference Sequence) haplotypes or CRS plus 16519C, which is considered a mutation hot spot (Brandstatter et al., 2003) were classified as H2, the haplogroup which the CRS belongs to. Haplogroup frequencies for both regions and areas within each region are reported in Table 4.2.1a.

**Table 4.2.1a**. Mitochondrial DNA haplogroup frequencies. In bold basal haplogroups in italic sub-branches.

| Haplogroup | **EUBOEA (95)** | C-Euboea (37) | SC-Euboea (35) | N-Euboea (23) | **KORINTHIA (86)** | NE-Korinthia (31) | S-Korinthia (21) | NW-Korinthia (34) |
|---|---|---|---|---|---|---|---|---|
| **F** | | | | | **0,012** | | **0,048** | |
| **H** | **0,295** | **0,243** | **0,371** | **0,261** | **0,430** | **0,323** | **0,524** | **0,471** |
| *H** | *0,032* | *0,054* | *0,029* | | | | | |
| *H1* | *0,032* | | *0,057* | *0,043* | *0,070* | *0,097* | *0,048* | *0,059* |
| *H2* | *0,179* | *0,135* | *0,229* | *0,174* | *0,279* | *0,194* | *0,381* | *0,294* |
| *H5* | *0,021* | *0,027* | | *0,043* | | | | |
| *H6* | *0,021* | *0,027* | *0,029* | | *0,035* | | *0,048* | *0,059* |
| *H7* | | | | | *0,023* | | *0,048* | *0,029* |
| *H8* | | | | | *0,012* | | | *0,029* |
| *H9* | | | | | *0,012* | *0,032* | | |
| *H20* | *0,011* | | *0,029* | | | | | |
| **HV** | **0,053** | **0,081** | **0,029** | **0,043** | **0,081** | **0,065** | **0,143** | **0,059** |
| **I** | | | | | **0,012** | | **0,048** | |
| **JT** | | | | | **0,012** | **0,032** | | |
| **J** | **0,137** | **0,162** | **0,029** | **0,261** | **0,070** | **0,097** | **0,000** | **0,088** |
| *J** | *0,063* | *0,054* | | *0,174* | *0,023* | | | *0,059* |
| *J1* | *0,074* | *0,108* | *0,029* | *0,087* | *0,047* | *0,097* | | *0,029* |
| **K** | **0,021** | **0,027** | **0,029** | **0,000** | **0,058** | **0,000** | **0,095** | **0,088** |
| *K** | *0,011* | | *0,029* | | *0,035* | | *0,048* | *0,059* |
| *K1* | *0,011* | *0,027* | | | *0,023* | | *0,048* | *0,029* |
| **M** | **0,021** | **0,054** | | | **0,035** | **0,065** | | **0,029** |
| **N1b** | **0,021** | **0,027** | | **0,043** | | | | |
| **P** | **0,021** | **0,027** | **0,029** | | | | | |
| **R** | **0,053** | **0,054** | **0,029** | **0,087** | **0,023** | **0,032** | | **0,029** |
| **T** | **0,084** | **0,108** | **0,057** | **0,087** | **0,081** | **0,065** | **0,095** | **0,088** |
| *T** | *0,011* | *0,027* | | | | | | |
| *T1* | *0,053* | *0,027* | *0,057* | *0,087* | | | | |
| *T2* | *0,021* | *0,054* | | | *0,081* | *0,065* | *0,095* | *0,088* |
| **U** | **0,253** | **0,189** | **0,343** | **0,217** | **0,174** | **0,323** | **0,048** | **0,118** |
| *U1* | *0,042* | *0,081* | *0,029* | | *0,035* | *0,097* | | |
| *U2* | *0,011* | | *0,029* | | | | | |
| *U3* | *0,021* | | *0,057* | | *0,012* | | *0,048* | |
| *U4* | *0,011* | *0,027* | | | *0,023* | *0,032* | | *0,029* |
| *U5* | *0,137* | *0,054* | *0,200* | *0,174* | *0,093* | *0,194* | | *0,059* |
| *U6* | *0,011* | | *0,029* | | | | | |
| *U7* | *0,011* | | | *0,043* | | | | |
| *U8* | *0,011* | *0,027* | | | *0,012* | | | *0,029* |
| **X** | **0,021** | **0,027** | **0,029** | | | | | |
| **Other** | **0,021** | | **0,057** | | **0,012** | | | **0,029** |

As for the wider European population, the most represented haplogroup in both Greek regions is H. The H2 haplogroup, the most frequent lineage within H, shows frequency which ranges from 52,4% to 24,3% (Southern Korinthia and Central Euboea respectively) whereas for the pooled samples the higher frequency has been observed in Korinthia (43%; Euboea=29,5%). The second most frequent haplogroup is U, and more precisely the U5 lineage which account for more than half of the U lineages for both Euboea and Korinthia. Interestingly for the Southern Korinthia sample U5 is virtually absent which could be the result of the incorrect haplogroup prediction and/or limited sampling. The haplogroup J exhibits higher frequencies in Euboea (13,7%) than in Korinthia (7%). The most representative branch of haplogroup J is J1 which accounts for more than half of all the J haplotypes. It is worth noting that no sample, within the J haplogroup, carried the 16231C mutation which is the defining mutation of the J2a1a sub-branch, considered as one of the lineages associated with Neolithic transition in Europe together with haplogroup K2a (Soares et al., 2010). The frequency of K haplogroup ranges between 2,1% and 5,8% with the higher value observed in Korinthia. The T haplogroup frequency is similar for both regions (8,4% and 8,1% for Euboea and Korinthia respectively). Other minor haplogroups (F, HV, I, JT, M, N1b, P, R and X) accounts for rest of the diversity which will not be further described.

The haplogroup structure of the Greek populations analysed resulted in accordance with previously published data (Richards et al., 2000) for Greek population. No further analysis based on these results has been carried out since the haplogroup classification is the result of a software prediction and is not currently reliable.

Internal diversity parameters for both haplogroups and haplotypes are reported in Table 4.2.1b. They show a higher diversity for both haplogroups and haplotypes in Euboea rather than Korinthia. If the single areas are taken into consideration, the haplogroup diversity ranges from 0,958 to 0,848 (Central Euboea and Southern Korinthia respectively) while the haplotype diversity from 0,992 to 0,959 (South-central and northern Euboea and North-western Korinthia respectively), in both cases the higher values are observed for Euboea areas. The MNPD also points to a lower diversity among the Korinthia areas and Korinthia region itself if compared with Euboea.

**Table 4.2.1b**. Intra-population diversity indices for MtDNA data. Abbreviations MNPD=Mean Number of Pairwise Differences, S= number of polymorphic sites, πn= nucleotide diversity.

| | N | N haplogroups | Haplogroup diversity | S | N haplotypes | Haplotype diversity | MNPD | Πn |
|---|---|---|---|---|---|---|---|---|
| **Euboea** | **95** | **28** | **0,932 (+/-0,013)** | **70** | **71** | **0,991 (+/-0,004)** | **5,460 (+/-2,651)** | **0,010 (+/-0,005)** |
| C-Euboea | 37 | 20 | 0,958 (+/-0,015) | 46 | 30 | 0,986 (+/-0,010) | 5,149 (+/-2,552) | 0,009 (+/-0,005) |
| SC-Euboea | 35 | 18 | 0,911 (+/-0,032) | 47 | 31 | 0,992 (+/-0,010) | 5,592 (+/-2,750) | 0,010 (+/-0,006) |
| N-Euboea | 23 | 11 | 0,917 (+/-0,031) | 34 | 21 | 0,992 (+/-0,015) | 5,589 (+/-2,785) | 0,010 (+/-0,006) |
| **Korinrhia** | **86** | **23** | **0,895 (+/-0,023)** | **70** | **57** | **0,969 (+/-0,012)** | **4,685 (+/-2,318)** | **0,009 (+/-0,005)** |
| NE-Korinthia | 31 | 12 | 0,910 (+/-0,027) | 35 | 25 | 0,985 (+/-0,012) | 4,755 (+/-2,389) | 0,009 (+/-0,005) |
| NW-Korinthia | 34 | 17 | 0,904 (+/-0,040) | 47 | 26 | 0,959 (+/-0,260) | 4,870 (+/-2,434) | 0,009 (+/-0,005) |
| S-Korinthia | 21 | 11 | 0,848 (+/-0,070) | 27 | 17 | 0,967 (+/-0,030) | 4,181 (+/-2,163) | 0,008 (+/-0,004) |

Through the analysis of mismatch distribution we detected signature of demographic expansion since all samples present bell-shaped distributions with low raggedness values which range from 0,008 to 0,015 (Figure 4.2.1a).

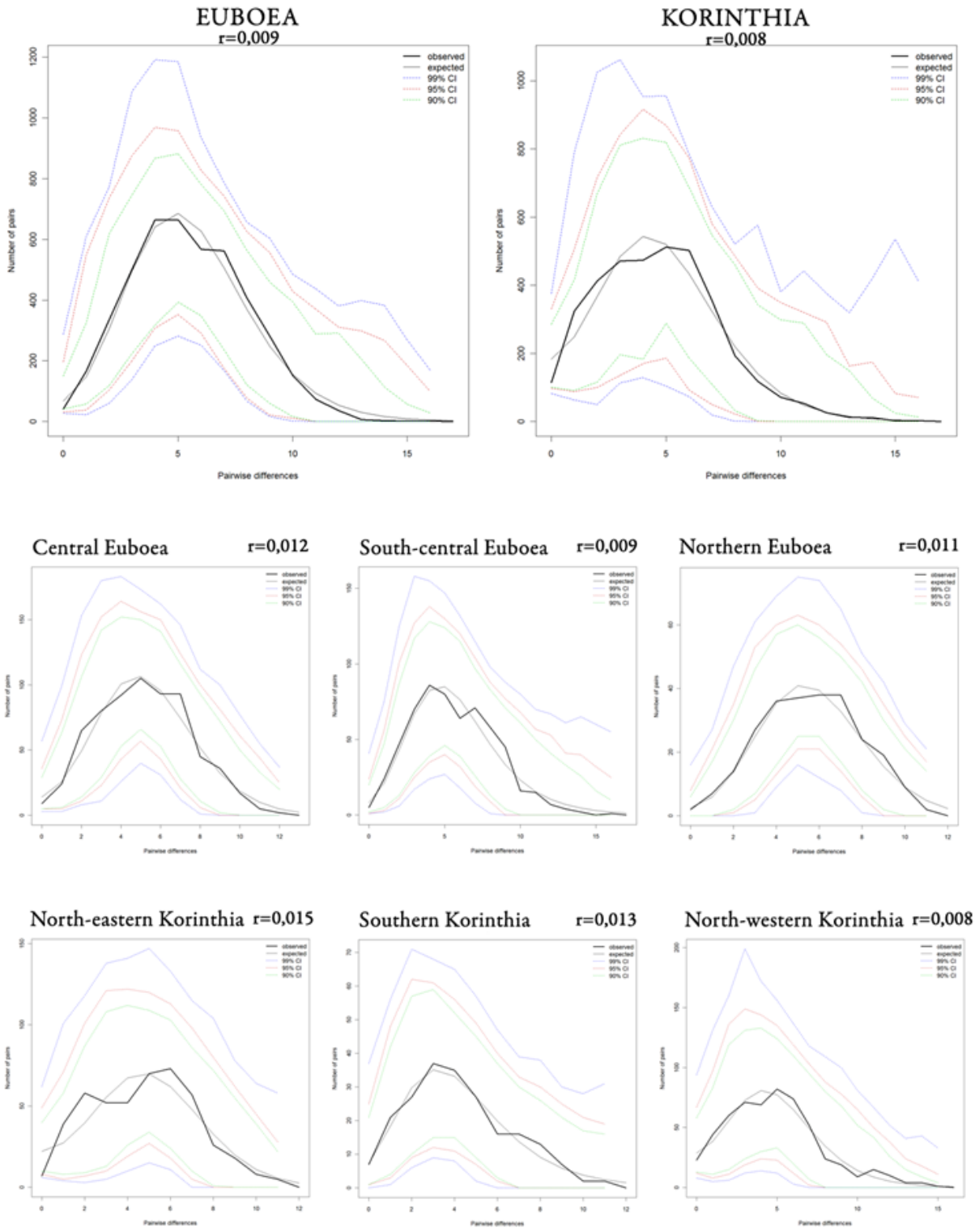



**Figure 4.2.1a**. Mismatch distributions of the populations analysed, r stands for the Raggedness.

### 4.2.2 Demographic inferences

The τ parameter, which is directly correlated to the mean and variance of the mismatch distribution, has been used to estimate the time of the demographic expansion through the simple formula τ=2μt (Schneider and Excoffier 1999). Confidence intervals have been obtained by a parametric bootstrap approach; the analysis has been performed with the Arlequin 3.5 package. Estimates of the time since demographic expansion obtained with this method are reported in Table 4.2.1c.

**Table 4.2.1c**. Estimates of the time since expansion (T) obtained from the τ parameter.

| | T | T (2,5% c.i.) | T (97,5% c.i.) |
|---|---|---|---|
| **Euboea** | **25419** | **14718** | **43356** |
| C-Euboea | 27022 | 14829 | 41397 |
| SC-Euboea | 22244 | 15068 | 45104 |
| N-Euboea | 32412 | 20246 | 40354 |
| **Korinthia** | **22915** | **11405** | **61138** |
| NE-Korinthia | 29387 | 16333 | 41980 |
| NW-Korinthia | 22288 | 13691 | 45831 |
| S-Korinthia | 14918 | 7470 | 36247 |

The values point to a pre-Neolithic demographic expansion for all populations with estimates that range from approximately 15Kya to 32Kya.

As for the Y chromosome, a coalescent approach for the inference of population demographic parameters was used. The T.M.R.C.A. values are very similar for all populations and range from approximately 43Kya to 47Kya, pointing to the first European peopling event which followed the Out of Africa. Moreover, the current effective population sizes resulted highly different between the two Greek regions with Euboea exhibiting the highest value. Posterior estimates of demographic parameter values are reported in Table 4.2.2a.

**Table 4.2.2a**. Posterior estimates of demographic parameters values obtained for MtDna. Abbreviations Ne= current effective population size**,** r= population growth rate and T.M.R.C.A.= Time of the most recent common ancestor.

| | Ne | Ne (95%c.i.) | r | r (95%c.i.) | T.M.R.C.A | T.M.R.C.A (95%c.i.) |
|---|---|---|---|---|---|---|
| **Euboea** | | | | | | |
| *mean* | 151831 | | $3,83x10^{-3}$ | | 46616 | |
| *median* | 142227 | 60024-282188 | $3,77x10^{-3}$ | $2,32x10^{-3}$-$5,69x10^{-3}$ | 45552 | 32491-63305 |
| *modal* | 119761 | | $3,69x10^{-3}$ | | 43795 | |
| **Korinthia** | | | | | | |
| *mean* | 95439 | | $3,44x10^{-3}$ | | 47533 | |
| *median* | 81058 | 32549-240826 | $3,34x10^{-3}$ | $1,80x10^{-3}$-$5,64x10^{-3}$ | 46232 | 32016-70521 |
| *modal* | 64795 | | $3,16x10^{-3}$ | | 44383 | |

The Bayesian skyline plots are presented in Figure 4.2.1b. For both regions the plot indicates a pre-Neolithic demographic growth in accordance with result obtained in a wider European scale (N=50) based on coding region sequences (Atkinson et al., 2008). Around 10Kya, consistent with the beginning of Neolithic in Greece, the two plots display a different profile. Euboea shows a slight increase in growth rate while for Korinthia no appreciable growth rate changes can be observed. Interestingly the Neolithic expansion detected for Euboea have not been seen in the wider European sample analysed in Atkinson et al., 2008. As reported in the article the more recent demographic events, like the Neolithic population growth, should be considered very carefully due to the limits of the temporal resolution of the method (Atkinson et al., 2008). However, it would be interesting to test whether using a larger number of individuals, possibly with a much lower resolution, from a well defined population sample, rather than a pool of European could result in more reliable demographic estimate.

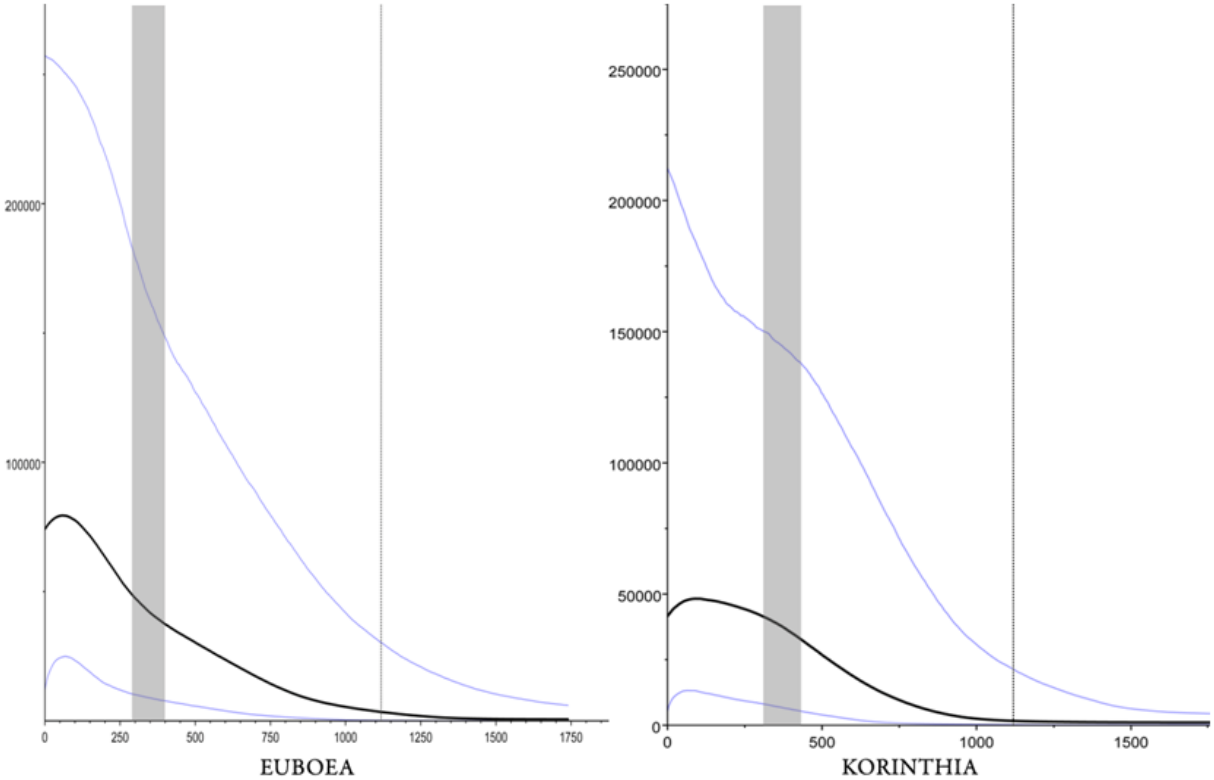



**Figure 4.2.2b**. Bayesian Skyline Plots of effective population size through time. Gray lines bound the 95% HPD for effective population size, accounting for uncertainty in the reconstructed phylogeny and substitution model parameters. Time is expressed in generations. Light gray block indicates a time range of 2Ky, from 10Kya to 8Kya considered consistent with the beginning of the Neolithic period in Greece.

### 4.2.3 Inter-population diversity

The mitochondrial DNA diversity between the two Greek regions and other European, Anatolian and Middle-eastern populations has been explored. A database of 32 populations has been assembled using literature data. To make the resolution all the data retrieved uniform, a 351 base pairs fragment of the HVR-I region has been considered. The list of population used for comparisons together with sample sizes and references are reported in Table 4.2.3a.

**Table 4.2.3a**. Population database used for comparisons. Insertions, deletions were not taken into consideration. Mutations in positions 16182 and 16183, when mutation at position 16189 was present, have not been considered.

| Population | Abbreviation | N | Reference |
|---|---|---|---|
| Croatians | CRO | 96 | Babalini et al., 2005 |
| Italians (Puglia) | PUG | 25 | Babalini et al., 2005 |
| Italians (Campania) | CMP | 47 | Babalini et al., 2005 |
| Italians (Lazio) | LAZ | 51 | Babalini et al., 2005 |
| Italians (Abbruzzo/Molise) | ABZ | 67 | Babalini et al., 2005 |
| Albanians | ALB | 41 | Belledi et al., 2000 |
| Italians (Bologna) | BOL | 100 | Bini et al.,2003 |
| Albanians | AL2 | 42 | Bosch et al., 2006 |
| Greeks (Thrace) | GR2 | 25 | Bosch et al., 2006 |
| Macedonians | MC2 | 37 | Bosch et al., 2006 |
| Bulgarians | BUL | 30 | Calafell et al., 1996 |
| Turks | TUR | 96 | Calafell et al., 1996; Comas et al., 1996; Richards et al., 2000 |
| Kurds | KUR | 29 | Comas et al., 2000 |
| Italians (Tuscany) | TUS | 49 | Francalacci et al., 1996 |
| Jordanians | JOR | 108 | Gonzalez et al., 2008 |
| Greeks (North) | GRE | 297 | Irwing et al., 2008 |
| Bosnians | BOS | 134 | Malyarchuk et al., 2003 |
| Slovenians | SLV | 101 | Malyarchuk et al., 2003 |
| Greeks (Crete) | CRT | 180 | Martinez et al., 2007 |
| Armenians | ARM | 42 | Nasidze & Stoneking 2001 |
| Georgians | GEO | 102 | Nasidze & Stoneking 2001 |
| Italians (Basilicata) | BSC | 92 | Ottoni et al., 2009 |
| Italians (Calabria) | CLB | 95 | Ottoni et al., 2009 |
| Italians (Sicily) | SLC | 154 | Ottoni et al., 2009 |
| Romanians | ROM | 88 | Richards et al., 2000 |
| Bulgarians | BLG | 110 | Richards et al., 2000 |
| Iraqi | IRQ | 116 | Richards et al., 2000 |
| Palestinians | PAL | 117 | Richards et al., 2000 |
| Syrians | SYR | 69 | Richards et al., 2000 |
| Greeks (Athens) | GRr | 65 | Richards et al., 2000 |
| Italians (Tocco) | TOC | 50 | Verginelli et al., 2003 |
| Macedonians | MAC | 182 | Zimmermann et. al., 2007 |

To investigate the genetic relationships between Euboea, Korinthia and the populations from the database a MDS plot based on the Fst genetic distance matrix has been drawn (Figure 4.2.3a.)

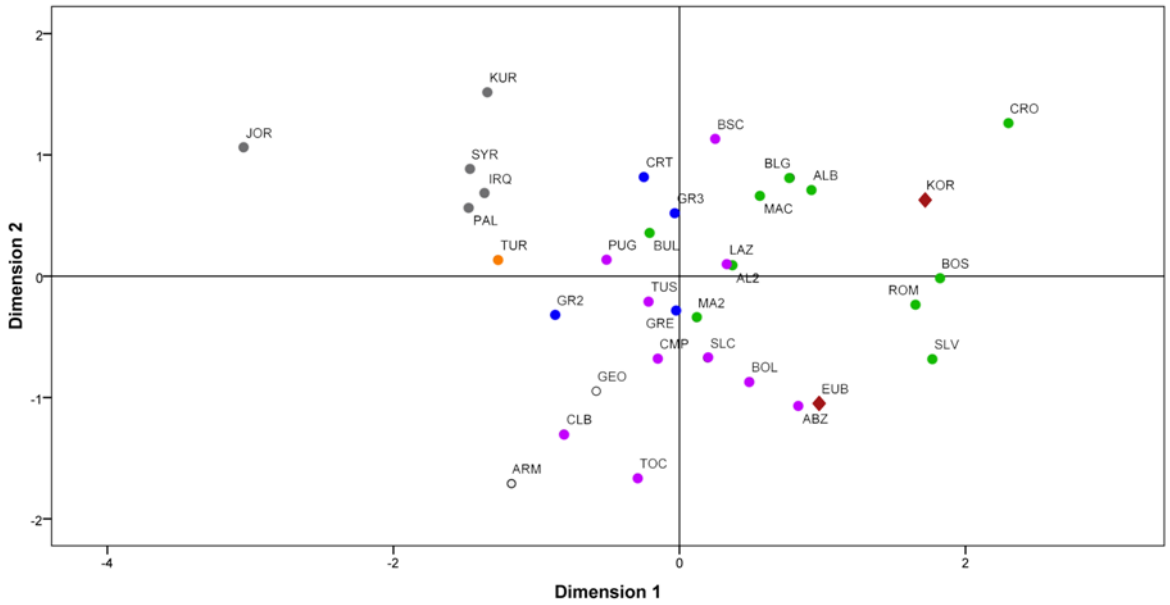



**Figure 4.2.3a**. Multidimentional Scaling plot based on Fst genetic distances. The stress value of 0,273 is acceptable according to Sturrocks and Rocha (2000). For population abbreviation see Table 5.3.2a. Statistically insignificant values were converted to 0.

Close affinities of the Anatolian (orange circle) and Middle-eastern populations (gray circles) are summarized by the plot. The European populations (blue circles= Greek, green circles= Balkan, purple circles= Italian) tend to cluster together. Interestingly the Greek populations closer to the Anatolia/Middle-east group are Crete (CRT) and the sample from the extreme northern region of Greece, Thrace (GR2). The two Greek populations analysed in this thesis, Euboea and Korinthia, fall close to the Italian populations and Balkan ones, respectively, and far away from the other Greek populations and the Anatolia/Middle-east group. The bidimensional representation of the genetic distances does not provide a clear indication on which of the two Greek regions is genetically closer to the Anatolia/Middle-east group. This information may be obtained inspecting the values of the Fst distances (Table 4.2.3b) which is more clearly highlighted by the third dimension of the MDS plot (data not shown). The lower genetic distances from the Anatolian and Middle-eastern populations are obtained for the Euboean population, although the differences are slight. More in details, the genetic distance value between Euboea and Turkey is approximately half than that obtained from Korinthia, and even lower than the distance between the two Greek regions. The only exception to this pattern is represented by the Syrian population which has a lower genetic distance with Korinthia. These observations indicate the slightly lower genetic distance of Euboea, rather than Korinthia, to the Anatolia/Middle-east group.

**Table 4.3.2b**. Matrix of genetic distances between the two Greek populations analysed in this thesis and the Anatolian and Middle-eastern ones.

| | EUB | KOR | TUR | IRQ | JOR | PAL | SYR |
|---|---|---|---|---|---|---|---|
| EUB | 0 | | | | | | |
| KOR | 0,0091 | 0 | | | | | |
| TUR | 0,0086 | 0,0148 | 0 | | | | |
| IRQ | 0,0119 | 0,0129 | 0,0044 | 0 | | | |
| JOR | 0,0263 | 0,0281 | 0,0082 | 0,0080 | 0 | | |
| PAL | 0,0155 | 0,0184 | 0 | 0 | 0,0101 | 0 | |
| SYR | 0,0144 | 0,0131 | 0 | 0 | 0,0069 | 0 | 0 |

A haplotype sharing analysis has been performed in order to quantify the genetic affinities of the two Greek regions with the other populations of the database (Figure 4.3.2b)

The analysis does not detect any clear geographical pattern. As expected, the European populations share more haplotypes with Euboea and Korinthia than Asian. Both these regions share the highest number of different haplotypes with the Northern Greek sample and the Macedonian sample, while the lowest number of shared different haplotypes, for both regions, is with the sample from Puglia. The other Italian populations share more haplotypes with Korinthia than with Euboea which, on the other hand, exhibits higher number of shared haplotypes with the Anatolia/Middle-east populations than Korinthia. If the frequency of the shared haplotypes is taken into consideration (dotted lines in Figure 4.3.2b) the pattern observed for the comparison of the two Greek populations and the Anatolian/Middle-east group seem to be inverted. In fact, Korinthia has a higher frequency of shared haplotypes with all the Middle-eastern populations than Euboea which, on the other hand, exhibits higher frequencies of sharing only with Turkey. Although, this higher frequencies of sharing seen for Korinthia is to be attributed mainly to CRS sequences which occur at a 21%, two times higher than Euboea. If the CRS sequences are excluded from the analysis the sharing frequencies of Korinthia drops sensibly to values lower than the ones observed for Euboea for all comparisons with the exception of the Syrian population (data not shown). The CRS sequences most probably belong to the H2 haplogroup which, until now, has never been associated with the spread of Neolithic technologies.

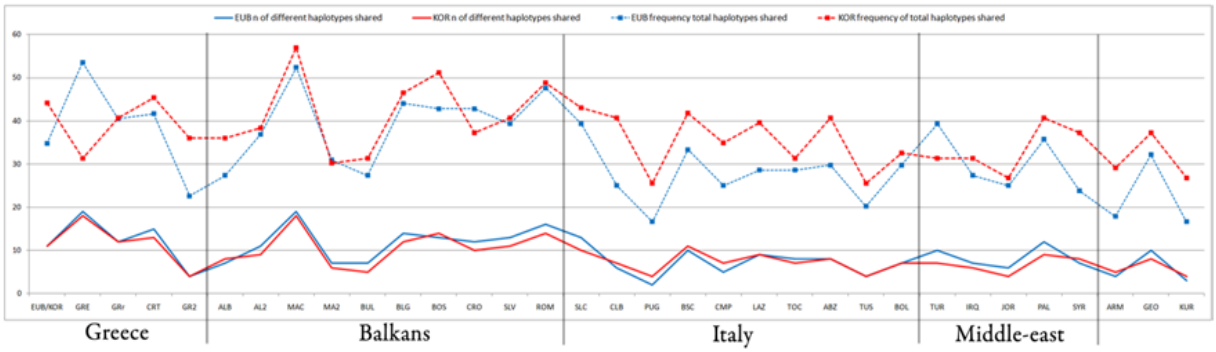



**Figure 4.3.2b**. Haplotype sharing analysis. The thick lines represent the number of different haplotypes shared between the two Greek regions (blue for Euboea and red for Korinthia) and all the other population. The dotted lines represent the frequency of shared haplotypes of each of the two Greek regions with all the other populations. For population abbreviation see Table 5.3.2a.

# 5. DISCUSSION

## 5.1 The Neolithic signature in Greece

As a contribution to the knowledge of the Neolithic migration into Greece and identification of the route followed by Neolithic farmers during the colonization process, the present study surveys two Greek regions supposed to be the arrival points of the two migration routes which have so far regarded to be the most likely. The two scenarios have clear expectations which were tested using evidence from both unilinear markers (Y-chromosome and MtDNA).

### 5.1.1 A male perspective

The results of our Y-chromosome survey provide a means to compare and contrast the role of migration in the establishment of the first Neolithic settlements in Greece.

As assessed earlier in this thesis, a higher genetic affinity of one or the other of the two populations analysed with Anatolian and Middle-eastern population should provide indications on which route have been followed during the Neolithic colonization of Greece. Regarding this point, a substantial agreement between PCA based on haplogroup frequencies and MDS based on genetic distances calculated for STRs indicates Euboea as the population which presents a closer genetic affinity with Anatolia and Middle-east. This conclusion is enforced by the vector analysis performed for the PCA plot. This latter analysis indicates that the Anatolia/Middle-eastern cluster is mainly associated with the J2a-M410 haplogroup which is often associated with Neolithic transition in Europe (Sengupta et al., 2006; King et al., 2008; Battaglia et al., 2009). Moreover, the PCA plot indicates a close relationship between Euboea, Crete and the Mediterranean region of Turkey (T6). The genetic data thus appears to support the long-held theory that maritime colonization of Crete and that the colonists that arrived in Mainland Greece probably came from this Turkey region (Evans 1921). In fact, some well known Neolithic sites, like Mersin/Yumuktepe and Tarsus, are located in this specific region. On the other hand, the second Greek region analysed seems to be more closely related with Balkan and other Greek populations. The MDS plot also suggests closer relationship of Euboea with Anatolian and Middle-eastern populations whereas Korinthia behaves as an outlier.

It is reasonable to think that some historical events that followed the Neolithic colonization which had as main actors the Greeks and the Anatolian populations could

have altered the haplogroup composition and diversity of the supposed Neolithic lineages. The 400 years occupation of Greece by the Ottoman Empire could have brought Anatolian lineages into Greece, through admixture events. Although, historical records report of intense settle of Ottoman populations only in the northern regions of Greece, especially Thrace (Treadgold 1997). According to these records admixture between Ottoman and Greek populations should have been very limited in the rest of Greece. Therefore, the signals captured by the PCA and the MDS analysis should mainly reflect more ancient events rather modern ones.

In order to gain insights in the past demographic dynamics of the two Greek populations analysed a Bayesian coalescent approach was used using Y-chromosomes STR data. The adoption of agriculture should have favoured a rapid and strong demographic expansion which probably has left a detectable genetic footprint (Ammerman & Cavalli-Sforza 1984; Boyle & Renfrew 2000). Estimates of time since expansion points towards an earlier adoption of agriculture in Euboea, approximately 8,8Kya whereas for Korinthia the estimate was slightly more recent (7Kya) dating more closely to the late Neolithic horizon. The same Bayesian approach was also used to estimate the times since expansion of the main Greek haplogroups in order to assess the haplogroups that could be associated with the Neolithic transition in Greece. The only haplogroup that showed times since expansion compatible with the Neolithic is J2a-M410 which has already been associated with Neolithic transition by several other studies (Sengupta et al., 2006; King et al., 2008; Battaglia et al., 2009). All the other haplogroups shows expansion dates consistent with the late Neolithic/initial Bronze Age horizon (R1b-M269 and E1b-V13) and the initial/late Bronze Age horizon (R1a-M17, G2a-P15, I2-M438, J1-M267 and J2b-M102). Interestingly, all the haplogroups that had previously considered as associated with Neolithic transition, namely R1b-M269 (Balaresque et al., 2009) and G2a-P15 (Cinnioglu et al., 2004; Battaglia et al., 2009) were not confirmed as such by our estimates. The oldest estimate of time since expansion of J2a-M410 was detected for Euboea (approximately 10Kya). Korinthia estimates were slightly more recent (approximately 9,4Kya). Nevertheless, the time span of ~600 years which separates Euboea and Korinthia is compatible with the lower bound of the supposed speed of spread of farming economy (0,6-1 Kmy) estimated by Pinhasi et al., 2005. The earlier presence of haplogroup J2a-M410 in Euboea and the time span between the latter one and Korinthia could suggest an initial arrival of Neolithic

migrants in this region and the subsequent spread south-eastward towards the Peloponnesus peninsula.

The haplogroup J2a-M410 exhibits the highest frequency in Euboea and especially in the two central regions reaching the frequency of 20% for South-central Euboea. As proposed by Barbujani (2000) geographic origins of haplogroup expansions can be inferred from both frequency and associated diversity, with spatial levels of accumulated microsatellite diversity providing a metric for assessing directionality of movement and to help disentangle complexities associated with population stratification. In order to examine the distribution patterns of both frequency and microsatellite diversity into Greece contour maps have been constructed. The geographical distribution of haplogroup J2a-M410 frequency points to two distinct centres of high values: Crete and Euboea. Despite the highest frequencies of J2a-M410 are found in Crete, the microsatellite diversity resulted higher in Euboea, especially in the central areas. Moreover, even when the Cretan J2a-M319 haplotypes, which due to their outlier microsatellite pattern were not been considered (Malaspina et al., 2001; Martinez et al., 2007), the STR diversity pattern do not changes significantly since the mean STR variance value of Crete does not exceed the one observed for the central areas of Euboea. In conclusion the frequency and STR diversity distributions points to Central Euboea as the possible geographical origin of dispersal of haplogroup J2a-M410 in Greece. The phylogenetic relationship, within haplogroup J2a-M410, of the two Greek regions analysed with other Greek, Anatolian and Middle-eastern population has been examined through the construction of a network. The clear simple star-like shape emerging from the network is indicative of an expansion from a single source. The most represented population in the central node was the Middle-eastern one which can be reasonably considered as the geographical origin of this expansion. As for the two Greek regions, Euboea shows a higher amount of sharing, within the central node and his single-step mutation neighbours, with Middle-eastern and Anatolian populations suggesting a closer genetic relationship with this populations regarding the J2a-M410 lineage.

The sub-structure of the J2a-M410 lineage does not reveal significant differences between the two Greek regions. This observation could be consistent with a model of rapid and intense expansion of this lineage from one source. In fact, if the expansion had been slower or less intense, stochastic or selective evolutionary processes could have

altered the observed sub-structure of this haplogroup generating an observable divergence between the source and the receiver population.

### 5.1.2 A female perspective

The mitochondrial DNA HVR-I region analysed in this thesis does not provide enough information to allow a correct haplogroup classification therefore only analyses on population level has been carried out.

In order to investigate the genetic relationship of the two Greek regions and their neighbouring European, Anatolian and Middle-eastern populations a MDS plot has been constructed. The plot highlights a closer genetic distance of the Anatolian/Middle-eastern cluster with Crete and the Greek population of Thrace and Macedonia. This latter result is substantially in accordance with the evidence coming from ancient MtDNA data that point to an inland migration of Neolithic farmers from Turkey, through northern Greece and the Balkans, towards central Europe (Haak et al., 2010). As for the two Greek populations analysed, both fall within the wider southern European diversity. Euboea falls close to some Italian populations, while the other Greek population, Korinthia tends to cluster with southern Balkan populations. If we focus our attention only on the genetic relationship of Euboea and Korinthia with the Anatolian/Middle-eastern group the Fst values provide evidence of a closer genetic distance of Euboea with this group rather than Korinthia, especially for the Turkish sample. The same suggestion is provided by the haplotype sharing. Euboea shares more different haplotypes and with higher frequencies with Turkey and Middle-eastern population indicating higher genetic affinities between them.

Demographic inferences for MtDNA suggest a pre-Neolithic demographic expansion for both Euboea and Korinthia dating to around 25Kya. Similar results are also highlighted by the Bayesian skyline plot (BSP). Moreover, the BSP also detects a slight increase in growth rate for Euboea which dates around 9Kya, consistent with the early Neolithic horizon. Despite this result should be interpreted very carefully, due to the low information that the only HVR-I can give, the signal clearly highlights a stronger demographic impact in Euboea than Korinthia. Surely a larger population sample and/or more genetic information provided by the HVR-II region or, even better, the coding region sequence, should give enough information to confirm, with a higher confidence, these results.

### 5.2 Final considerations

It is generally accepted that Y-chromosome and mtDNA shows extremely different distribution patterns of variation in Europe (Rosser 2000). Y chromosome shows strong geographical structuring whereas, mtDNA reveals a more homogeneous landscape, at least at low resolution. This different pattern could be due probably to a higher level of female mobility among the European populations. Moreover, all Europeans essentially share the same set of mitochondrial lineages as Middle-easterners (Torroni et al., 2006), making any attempt to distinguish different migration events extremely difficult.

Our data seem to reflect this background, since the two genetic markers does not always provide similar results. For instance, the two markers tend to be concordant for Korinthia when compared with other European and Middle-eastern populations, they produce different patterns for Euboea. The difference between Y-chromosome and MtDNA data for this latter region has a potential explanation. As far as we know, demic diffusion involved both females and males, but a disparity between mtDNA and Y-chromosomal patterns could arise from an increased and transmitted reproductive success for male farmers compared to indigenous hunter-gatherers, without a corresponding difference between females (Balaresque et al., 2009). Demographic inferences of population growth, as far as they can be compared, do not exhibit significant differences. The growth rate values are similar for both markers in both populations. This could stand to indicate that the male and female portion of the populations have not undergone through differential demographic processes. Although, this latter observation does not exclude the possibility of a differential reproductive success of indigenous female hunter-gatherers and migrant female farmers. Moreover, the time frame of the slight increase in growth rate observed, through the BSP, for Euboea is in complete accordance with the times since expansion inferred from Y chromosome STR data. On the other hand, the lack of signals of increased growth rate for Korinthia could be tracing back to the differential reproductive success of female farmers, as posted earlier. Otherwise, the low resolution of mtDNA could be invoked. Unfortunately, the times since expansion for mtDNA and Y-chromosome are inferred with two extremely different methods that does not allow a direct comparison. In fact, while for Y-chromosome the estimates are calculated trough a Bayesian approach, which tends to capture the more intense expansion, for mtDNA the $\tau$ parameter was used

which, on the contrary, tends to capture the signals of the most ancient expansion (Schneider & Excoffier 1999).

It is also worth noting that the usage of single lineages, especially when investigating on past demographic dynamics, could be misleading. Several studies argue with the usage of single lineages for evolutionary inferences since the fluctuations in the effective population size could cause significant variations on haplogroup frequencies across generations (Zhivotovsky et al., 2006). This effect is more pronounced for both Y-chromosome and mtDNA since the effective population size of these markers is one fourth of that of any autosomal, which makes them more susceptible to stochastic processes. These frequency variations could depend on either stochastic or selective processes that could act as confounding factors in evolutionary inferences. Therefore, according to some authors (Pritchard et al., 1999) a population approach would be more informative on past demographic events.

In this thesis both approaches were used in the attempt to fully investigate the demographic dynamics of the two Greek populations.

# 6. CONCLUSIONS

In this thesis Y-chromosome and mtDNA data have been analysed in order to test two possible scenarios of colonization of Greece, from Middle-Eastern populations, during the early Neolithic.

As reported earlier, only one other recent study, tried to explore the impact of Neolithic farmers into Greece (King et al., 2008). This study pointed to the CCPr route as the most probable based only on Y-chromosome data, with a lower resolution, a more restricted geographical coverage of the Greek regions and a lower sample size, compared to the present research. Furthermore, the present study investigates both uniparental markers. Nonetheless, both population and single lineages approaches have been considered in order to exploit as much as possible the potential of our data. Finally, the usage of more sophisticated statistical analyses and also the investigation of demographic dynamics, which were not considered in King et al.'s work, could have hopefully led to more reliable results.

The two scenarios, the Island hopping route (IHr) and the Cyprus/Crete/NE-Peloponnesus route (CCPr), have been tested through the comparison of the theoretical expectations and the empirical observations. This comparison is summarized in Table 6a.

All observations point to same conclusion with the only exception of the population expansion times for mtDNA which does not provide enough information to fulfil the theoretical expectation. This is due mostly to the method used for the inferences which tends to capture the more ancient expansion rather than the most intense. Moreover, the slight expansion signals captured by the BSP have not been considered informative enough to answer this question, since the usage of the only HVRI region of the mtDNA, for this kind of analysis, have not yet achieved a strong theoretical supported.

On the whole, our results point strongly to the Island hopping route (6 out of 7 expectations) which brought Neolithic migrants, from the Middle-eastern Mediterranean coasts or more probably from the Anatolian Mediterranean area, to Euboea. The arrival of Neolithic farmers would have been followed by a rapid demographic and spatial expansion which allowed the spread of farming technologies to the rest of mainland Greece.

**Table 6a.** Expectations for the two migration routes compared to observations. IHr=Island hopping route; CCPr= Cyprus/Crete/NE Peloponnesus route.

| | EXPECTATIONS | OBSERVATIONS | IHr | CCPr |
|---|---|---|---|---|
| Y-CHROMOSOME | Genetic affinity of sampled populations with Anatolian/Middle Eastern populations | Closer genetic relationship of Euboea than Korinthia with the Middle-eastern and Anatolian populations as revealed by both PCA and MDS analyses | √ | |
| | Population expansion times | Older times since expansion detected for Euboea, compatible with an earlier Neolithic horizon. | √ | |
| | Frequency and variance of "Neolithic" haplogroups | Higher frequency and mean STR variance of haplogroup J2a-M410 in Euboea. Spatial distribution highlights a possible dispersal of this lineage from Euboea island rather than Korinthia | √ | |
| | Time since expansion for Neolithic Haplogroups | Earlier time since expansion estimates for the J2a-M410 Neolithic lineage in Euboea consistent with a demic diffusion. | √ | |
| | Genetic affinities with Anatolian/Middle-eastern population for Neolithic haplogroups | The J2a-M410 network shows a higher representation of Euboean haplotypes in both central node and in his single-step mutation neighbours, suggesting a higher genetic affinity within this haplogroup with Middle-eastern and Anatolian populations. | √ | |
| MTDNA | Genetic affinity of sampled populations with Anatolian/Middle Eastern populations | Euboea exhibits lower genetic distances from Anatolian and Middle-eastern populations than Korinthia. Haplotype sharing also points to a closer genetic affinity of Euboea with the Anatolian and Middle-eastern populations. | √ | |
| | Population expansion times | Both populations exhibit pre-Neolithic times since expansion. The Bayesian skyline plot detects a slight demographic increase within the early Neolithic time frame only for Euboea. | - | - |

Unfortunately, our conclusions are well supported by the Y-chromosome results whereas mtDNA does not provide so clear results. There are two ways, not mutually exclusive, in which this uncertainty could be overcome. Surely, a higher resolution of the latter genetic marker could provide further information, giving the possibility to explore the maternal Neolithic signatures in Greece in more detail. Moreover, a full geographical coverage of the Greek territory will surely contribute to a more complete comprehension of the genetic structure, of both uniparental markers, of Greek population, which could help in the identification of more recent migration either other evolutionary processes that, in the present study, could have acted as confounding factors that could have driven to erroneous conclusions.

In conclusion, this thesis provides new and, as we believe, strong evidence for the understanding of the migration processes that brought farming economy into Greece.

Hopefully, further investigation should be carried out in this direction. The usage of a genomic approach could be an important advance for this kind of studies, giving the possibility to bypass the limitations of the uniparentaly transmitted markers.

# REFERENCES


Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A, and Santachiara-Benerecetti AS. 2003. Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. Mol. Phylogenet. Evol 28:458-472.

Ammerman AJ and Cavalli-Sforza LL. 1973. A population model for the diffusion of early farming in Europe. in C Renfrew, (eds) The explanation of culture change. London: Duckworth.

Ammerman AJ and Cavalli-Sforza LL. 1984. The neolithic transition and the genetics of populations in Europe. Princeton N.J.: Princeton University Press.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, et al. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457-465.

Andreou S, Fotiadis M and Kotsakis K. 1996. Review of Aegean Prehistory V: the Neolithic and Bronze Age of Northern Greece. American Journal of Archaeology 100: 537–97.

Atkinson QD, Gray RD and Drummond AJ. 2008. mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. Mol. Biol. Evol 25:468-474.

Babalini C, Martínez-Labarga C, Tolk H, Kivisild T, Giampaolo R, Tarsi T, Contini I, Barać L, Janićijević B, Martinović Klarić I, et al. 2005. The population history of the Croatian linguistic minority of Molise (southern Italy): a maternal view. Eur. J. Hum. Genet 13:902-912.

Bagolini B, and Biagi P. 1990. The radiocarbon chronology of the neolithic and copper age of northern Italy. Oxford J Archeol 9:1-23.

Balaresque P, Bowden GR, Adams SM, Leung H, King TE, Rosser ZH, Goodwin J, Moisan J, Richard C, Millward A, et al. 2010. A predominantly neolithic origin for European paternal lineages. PLoS Biol 8:e1000285.

Bandelt HJ, Forster P, and Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol 16:37-48.

Barbujani G. 2000. Geographic patterns: how to identify them and why. Hum. Biol 72:133-153.

Battaglia V, Fornarino S, Al-Zahery N, Olivieri A, Pala M, Myres NM, King RJ, Rootsi S, Marjanovic D, Primorac D, et al. 2009. Y-chromosomal evidence of the cultural diffusion of agriculture in Southeast Europe. Eur. J. Hum. Genet 17:820-830.

Binder D. 2000 Mesolithic and Neolithic interaction in southern France and northern Italy: new data and current hypotheses, in TD Price (eds) Europe's first farmers. New York: Cambridge University Press

Bini C, Ceccardi S, Luiselli D, Ferri G, Pelotti S, Colalongo C, Falconi M, and Pappalardo G. 2003. Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). Forensic Sci. Int 135:48-52.

Bogucki P. 2009. Forest farmers and stockherders : early agriculture and its consequences in North-Central Europe. Cambridge: Cambridge University Press.

Bosch E, Calafell F, González-Neira A, Flaiz C, Mateu E, Scheil H, Huckenbeck W, Efremovska L, Mikerezi I, Xirotiris N, et al. 2006. Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. Ann. Hum. Genet 70:459-487.

Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, Antanaitis-Jacobs I, Haidle MN, Jankauskas R, Kind C, et al. 2009. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science 326:137-140.

Brandstätter A, and Parson W. 2003. Mitochondrial DNA heteroplasmy or artefacts-a matter of the amplification strategy? Int. J. Legal Med 117:180-184.

Brandstätter A, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W, and Parsons TJ. 2004. Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. Int. J. Legal Med 118:294-306.

Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, e Hammer MF. 2002. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. Forensic Sci. Int 129:10-24.

Calafell F, Underhill P, Tolun A, Angelicheva D, and Kalaydjieva L. 1996. From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. Ann. Hum. Genet 60:35-49.

Capelli C, Brisighelli F, Scarnicci F, Arredi B, Caglia' A, Vetrugno G, Tofanelli S, Onofri V, Tagliabracci A, Paoli G, et al. 2007. Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic-Neolithic encounter. Mol. Phylogenet. Evol 44:228-239.

Chapman J. 1991. Els origens de l'agricultura al Sud-Est d'Europa. Cota Zero 7:126–35.

Chapman J. 1993. Cultural transformations and interactions in Eastern Europe. Aldershot [England]; Brookfield USA: Avebury.

Childe VG. 2003. The dawn of European civilization. New ed. London: Kegan Paul International.

Childe VG. 1928 The Most Ancient East , London: Routledge and Kegan Paul.

Childe VG. 1934 New Light on the Most Ancient East , London: Kegan Paul, Trench, Trubner

Cinnioğlu C, King R, Kivisild T, Kalfoğlu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, et al. 2004. Excavating Y-chromosome haplotype strata in Anatolia. Hum. Genet 114:127-148.

Cohen M, and Wenner-Gren Foundation for Anthropological Research.;State University of New York College at Plattsburgh. 1984. Paleopathology at the origins of agriculture. New York: Academic Press.

Cohen M. 1977. The food crisis in prehistory : overpopulation and the origins of agriculture. New Haven [Conn.]: Yale University Press.

Cohen M. 1989. Health and the rise of civilization. New Haven: Yale University Press.

Colledge S, Conolly J, and Shennan S. 2004. Archaeobotanical Evidence for the Spread of Farming in the Eastern Mediterranean. CURR ANTHROPOL 45:S35-S58.

Comas D, Calafell F, Bendukidze N, Fañanás L, and Bertranpetit J. 2000. Georgian and kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. Am. J. Phys. Anthropol 112:5-16.

Comas D, Calafell F, Mateu E, Pérez-Lezaun A, and Bertranpetit J. 1996. Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. Mol. Biol. Evol 13:1067-1077.

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, et al. 2004. Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa. Am. J. Hum. Genet 74:1014-1022.

Cruciani F, La Fratta R, Torroni A, Underhill PA, and Scozzari R. 2006. Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): a posteriori evaluation of a microsatellite-network-based approach through six new biallelic markers. Hum. Mutat 27:831-832.

Cruciani F, La Fratta R, Trombetta B, Santolamazza P, Sellitto D, Colomb EB, Dugoujon J, Crivellaro F, Benincasa T, Pascone R, et al. 2007. Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. Mol. Biol. Evol 24:1300-1311.

Davis JL. 1992 Review of the Aegean Prehistory I: the islands of the Aegean. American Journal of Archaeology 96:699-756

Demoule J, and Perles C. 1993. The Greek Neolithic: A new review. J World Prehist 7:355-416.

Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, Brdicka R, Barbujani G, Papola F, Ciavarella G, et al. 2004. Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. Hum. Genet 115:357-371.

Donahue RE. 1992. Desperately seeking Ceres: A Critical Examination of Current Models for the Transition to Agriculture in Mediterranean Europe, in A.B.Gebauer and

T.D.Price (eds) Transitions to Agriculture in Prehistory , Madison: Prehistory Press, Monographs in World Archaeology 4.

Drummond AJ, Rambaut A, Shapiro B, and Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol 22:1185-1192.

Evans AJ. 1921. The Palace of Minos at Knossos. Volume I. London: Macmillan.

Excoffier L, and Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564-567. Kruskal JB. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29:1-27.

Flores C, Maca-Meyer N, Larruga JM, Cabrera VM, Karadsheh N, and Gonzalez AM. 2005. Isolates in a corridor of migrations: a high-resolution analysis of Y-chromosome variation in Jordan. J. Hum. Genet 50:435-441.

Francalacci P, Bertranpetit J, Calafell F, and Underhill PA. 1996. Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. Am. J. Phys. Anthropol 100:443-460.

González AM, Karadsheh N, Maca-Meyer N, Flores C, Cabrera VM, and Larruga JM. 2008. Mitochondrial DNA variation in Jordanians and their genetic relationship to other Middle East populations. Ann. Hum. Biol 35:212-231.

Greenfield, H and Jongsma, T. 2008 Sedentary pastoral gatherers in the Early Neolithic – architectural, botanical, and zoological evidence for mobile economies from Foeni-Salas, SW Romania, in W Douglass (eds) Living Well Together? Settlement and Materiality in the Neolithic of South-East and Central Europe. Oxford: Oxbow Books.

Gronenborn D. 1999 A variation on the basic theme: The transition to farming in southern central Europe. Journal of World Prehistory 2:23–210.

Gusmão L, Sánchez-Diz P, Calafell F, Martín P, Alonso CA, Alvarez-Fernández F, Alves C, Borjas-Fajardo L, Bozzo WR, Bravo ML, et al. 2005. Mutation rates at Y chromosome specific microsatellites. Hum. Mutat 26:520-528.

Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CSI, Brandt G, Schwarz C, Nicklisch N, et al. 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. PLoS Biol 8:e1000536.

Hall JM. 2000. Ethnic Identity in Greek Antiquity, Cambridge: Cambridge University Press.

Hammer MF, Blackmer F, Garrigan D, Nachman MW, and Wilder JA. 2003. Human population structure and its effects on sampling Y chromosome sequence variation. Genetics 164:1495-1509.

Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Santachiara-Benerecetti S, Oppenheim A, Jobling MA, Jenkins T, et al. 2000. Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. Proc. Natl. Acad. Sci. U.S.A 97:6769-6774.

Hanson EK, Berdos PN, and Ballantyne J. 2006. Testing and evaluation of 43 "noncore" Y chromosome markers for forensic casework applications. J. Forensic Sci 51:1298-1314.

Hillman GC, and Davies MS. 1990. Measured domestication rates in wild wheats and barley under primitive cultivation, and their archaeological implications. J World Prehist 4:157-222.

Irwin J, Saunier J, Strouss K, Paintner C, Diegoli T, Sturk K, Kovatsi L, Brandstätter A, Cariolou MA, Parson W, et al. 2008. Mitochondrial control region sequences from northern Greece and Greek Cypriots. Int. J. Legal Med 122:87-89.

Johnson M. 1996. Water, animals and agricultural technology: a study of settlement patterns and economic change in Neolithic southern Greece. Oxford J Archeol 15:267-295.

King RJ, Ozcan SS, Carter T, Kalfoğlu E, Atasoy S, Triantaphyllidis C, Kouvatsi A, Lin AA, Chow CT, Zhivotovsky LA, et al. 2008. Differential Y-chromosome Anatolian influences on the Greek and Cretan Neolithic. Ann. Hum. Genet 72:205-214.

Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, and Kronenberg F. 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum. Mutat. 32:25-32.

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, Roseman C, Underhill PA, Cavalli-Sforza LL, and Herrera RJ. 2004. The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Am. J. Hum. Genet 74:532-544.

MacNeish R. 1992. The origins of agriculture and settled life. 1° ed. Norman: University of Oklahoma Press.

Malaspina P, Ciminelli BM, Viggiano L, Jodice C, Cruciani F, Santolamazza P, Sellitto D, Scozzari R, Terrenato L, Rocchi M, et al. 1997. Characterization of a small family (CAIII) of microsatellite-containing sequences with X-Y homology. J. Mol. Evol 44:652-659.

Malaspina P, Tsopanomichalou M, Duman T, Stefan M, Silvestri A, Rinaldi B, Garcia O, Giparaki M, Plata E, Kozlov AI, et al. 2001. A multistep process for the dispersal of a Y chromosomal lineage in the Mediterranean area. Ann. Hum. Genet 65:339-349.

Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Drobnic K, and Miścicka-Sliwka D. 2003. Mitochondrial DNA variability in Bosnians and Slovenians. Ann. Hum. Genet 67:412-425.

Marjanovic D, Fornarino S, Montagna S, Primorac D, Hadziselimovic R, Vidovic S, Pojskic N, Battaglia V, Achilli A, Drobnic K, et al. 2005. The peopling of modern Bosnia-

Herzegovina: Y-chromosome haplogroups in the three main ethnic groups. Ann. Hum. Genet 69:757-763.

Martinez L, Underhill PA, Zhivotovsky LA, Gayden T, Moschonas NK, Chow CT, Conti S, Mamolini E, Cavalli-Sforza LL, and Herrera RJ. 2007. Paleolithic Y-haplogroup heritage predominates in a Cretan highland plateau. Eur. J. Hum. Genet 15:485-493.

Menozzi P, Piazza A, and Cavalli-Sforza L. 1978. Synthetic maps of human gene frequencies in Europeans. Science 201:786-792.

Morelli L, Contu D, Santoni F, Whalen MB, Francalacci P, and Cucca F. 2010. A comparison of Y-chromosome variation in Sardinia and Anatolia is more consistent with cultural rather than demic diffusion of agriculture. PLoS ONE 5:e10419.

Mountjoy PA. 1998. The East Aegean-West Anatolian Interface in the Late Bronze Age: Mycenaeans and the Kingdom of Ahhiyawa. Anatolian Studies 48:33.

Myres NM, Rootsi S, Lin AA, Järve M, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, Pshenichnov A, Yunusbayev B, et al. 2011. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur. J. Hum. Genet 19:95-101.

Nandris J. 1977. The perspective of long-term change in south-east Europe. In F Carter (ed.), An Historical Geography of the Balkans. New York: Academic Press

Nasidze I, and Stoneking M. 2001. Mitochondrial DNA variation and language replacements in the Caucasus. Proc. Biol. Sci 268:1197-1206.

Onofri V, Alessandrini F, Turchi C, Pesaresi M, Buscemi L, and Tagliabracci A. 2006. Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. Forensic Sci. Int 157:23-35.

Ottoni C, Martinez-Labarga C, Vitelli L, Scano G, Fabrini E, Contini I, Biondi G, and Rickards O. 2009. Human mitochondrial DNA variation in Southern Italy. Ann. Hum. Biol 36:785-811.

Ozdogan M. 1993. Vinca and Anatolia: a new look at a very old problem (or redefining Vinca culture from the perspective of Near Eastern tradition). In J Roodenberg (eds) Anatolia and the Balkans. Anatolica 19: 173–93.

Özdoğan M. 1997. The beginning of Neolithic economies in southeastern Europe: An Anatolian perspective. Journal of European Archeology 5:1–33.

Perlès C. 2001. The early Neolithic in Greece: The first farming communities in Europe. Cambridge; New York: Cambridge University Press.

Pinhasi R, Fort J, and Ammerman AJ. 2005. Tracing the origin and spread of agriculture in Europe. PLoS Biol 3:e410.

Pinhasi R. and Pluciennik M. 2004. A regional biological approach to the spread of farming in Europe: Anatolia, the Levant, South-Eastern Europe, and the Mediterranean. Current Anthropology 45:S59–S82.

Pluciennik M. 1998. Archaeology, archaeologists and 'Europe. Antiquity 278:816-824

Polunin O. 1987. Flowers of Greece and the Balkans : a field guide. 1° ed. Oxford: Oxford University Press.

Poncet P. 2009. Modeest: Mode Estimation. *R package version 1.09*.

Posada, and Crandall. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends Ecol. Evol. (Amst.) 16:37-45.

Price T. 2000. Europe's first farmers. New York: Cambridge University Press.

Radovanovic I. 2000. Houses and Burials at Lepenski Vir. European Journal of Archaeology 3:330-349.

Regueiro M, Cadenas AM, Gayden T, Underhill PA, and Herrera RJ. 2006. Iran: tricontinental nexus for Y-chromosome driven migration. Hum. Hered 61:132-143.

Renfrew C, and Aspinall A. 1990. Aegean obsidian and Franchthi Cave, in CPerles (eds) Les industries lithiques tailles de Franchthi (Argolide, Grece), Tome II. Les industries du mesolithique et neolithique initial, Excavations at Franchthi cave, Greece 5. Bloomington and Indianapolis: Indiana University Press.

Renfrew C, and McDonald Institute for Archaeological Research. 2000. Archaeogenetics : DNA and the population prehistory of Europe. Cambridge ;Oakville CT: McDonald Institute for Archaeological Research ;David Brown Book Co. [distributor].

Renfrew C. 1987 Archaeology and language : the puzzle of Indo-European origins, London: Pimlico.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. Am. J. Hum. Genet 67:1251-1276.

Rogers AR, and Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol 9:552-569.

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, et al. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am. J. Hum. Genet 67:1526-1543.

Runnels C. and Van Andel T.H. 1988. Trade and the Origins of Agriculture in the Eastern Mediterranean. Journal of Mediterranean Archaeology1.

Sanchez JJ, Hallenberg C, Børsting C, Hernandez A, and Morling N. 2005. High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males. Eur. J. Hum. Genet 13:856-866.

Schneider PM, Meuser S, Waiyawuth W, Seo Y, and Rittner C. 1998. Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. Forensic Sci. Int 97:61-70.

Schneider S, and Excoffier L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics 152:1079-1089.

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, et al. 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. Am. J. Hum. Genet 74:1023-1034.

Semino O, Passarino G, Quintana-Murci L, Liu A, Béres J, Czeizel A, and Santachiara-Benerecetti AS. 2000. MtDNA and Y chromosome polymorphisms in Hungary: inferences from the palaeolithic, neolithic and Uralic influences on the modern Hungarian gene pool. Eur. J. Hum. Genet 8:339-346.

Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CT, Lin AA, Mitra M, Sil SK, Ramesh A, et al. 2006. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. Am. J. Hum. Genet 78:202-221.

Shen P, Wang F, Underhill PA, Franco C, Yang WH, Roxas A, Sung R, Lin AA, Hyman RW, Vollrath D, et al. 2000. Population genetic implications from sequence variation in four Y chromosome genes. Proc. Natl. Acad. Sci. U.S.A 97:7354-7359.

Shi W, Ayub Q, Vermeulen M, Shao R, Zuniga S, van der Gaag K, de Knijff P, Kayser M, Xue Y, and Tyler-Smith C. 2010. A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. Mol. Biol. Evol 27:385-393.

Smith P, Bar-Yosef O, and Sillen A. 1984. Archaeological and Skeletal Evidence for Dietary Change During the Late Pleistocene/Early Holocene in the Levant, in MH Cohen and GJ Armelagos eds Palaeopathology at the Origins of Agriculture , New York: Academic Press.

Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt H, Torroni A, and Richards MB. 2010. The archaeogenetics of Europe. Curr. Biol 20:R174-183.

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, and Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. Am. J. Hum. Genet 84:740-759.

Sturrock K, and Rocha J. 2000. A Multidimensional Scaling Stress Evaluation Table. Field Methods 12:49-60.

Thorpe I. 1996. The origins of agriculture in Europe. London: Routledge.

Treadgold W. 1997. A history of the Byzantine state and society. Stanford (California): Stanford University Press.

Tringham R. 2000. Southeastern Europe in the transition to agriculture in Europe: bridge, buffer, or mosaic, in TD Price (eds) Europe's first farmers. New York: Cambridge University Press.

Underhill PA, and Kivisild T. 2007. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu. Rev. Genet 41:539-564.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazón Lahr M, Foley RA, Oefner PJ, and Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann. Hum. Genet 65:43-62.

Van Andel TH. 1996. Beyond the Acropolis : a rural Greek past. Stanford Calif.: Stanford University Press.

Van Andel TH, and Runnels CN. 1995. The earliest farmers in Europe. Antiquity 69:481–500.

Van Oven M, and Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. 30:E386-E394.

Verginelli F, Donati F, Coia V, Boschi I, Palmirotta R, Battista P, Mariani Costantini R, and Destro-Bisol G. 2003. Variation of the hypervariable region-1 of mitochondrial DNA in central-eastern Italy. J. Forensic Sci 48:443-444.

Whittle A. 1996. Europe in the Neolithic : the creation of new worlds. Cambridge;New York: Cambridge University Press.

Whittle, A. 1997. Moving on and moving around: Neolithic settlement mobility, in P Topping (eds) Neolithic Landscapes. Oxford: Oxbow Monograph 86.

Wijnen M. 1993 Early ceramics: local manufacture versus widespread distribution. Anatolica 19:319-31.

Wilson IJ, Weale ME, and Balding DJ. 2003. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. J Royal Statistical Soc A 166:155-188.

Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, and Goldstein DB. 2001. Genetic evidence for different male and female roles during cultural transitions in the British Isles. Proc. Natl. Acad. Sci. U.S.A 98:5078-5083.

Zalloua PA, Xue Y, Khalife J, Makhoul N, Debiane L, Platt DE, Royyuru AK, Herrera RJ, Hernanz DFS, Blue-Smith J, et al. 2008. Y-chromosomal diversity in Lebanon is structured by recent historical events. Am. J. Hum. Genet 82:873-882.

Zhivotovsky LA, Underhill PA, Cinnioğlu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, et al. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am. J. Hum. Genet 74:50-61.

Zhivotovsky LA, and Underhill PA. 2005. On the evolutionary mutation rate at Y-chromosome STRs: comments on paper by Di Giacomo et al. (2004). Hum. Genet 116:529-532.

Zhivotovsky LA, Underhill PA, e Feldman MW. 2006. Difference between evolutionarily effective and germ line mutation rate due to stochastically varying haplogroup size. Mol. Biol. Evol 23:2268-2270.

Zilhao J. 2000. From the Mesolithic to the Neolithic in the Iberian peninsula, in TD Price (eds) Europe's first farmers. New York: Cambridge University Press.

Zimmermann B, Brandstätter A, Duftner N, Niederwieser D, Spiroski M, Arsov T, and Parson W. 2007. Mitochondrial DNA control region population data from Macedonia. Forensic Sci Int Genet 1:e4-9.

Zvelebil M. 2000. The social context of the agricultural transition in Europe. In C Renfrew and K Boyle (eds) Archaeogenetics: DNA and the population prehistory of Europe. Cambridge ;Oakville CT: McDonald Institute for Archaeological Research ;David Brown Book Co. [distributor].

# Aknowledgments

I would like to thank Prof. Donata Luiselli and Prof. Davide Pettener for the opportunity to work with great intellectual freedom in this field, the patience they always showed, the confidence they have always shown and the significant contribution given for this thesis. Thanks also to all the other people of the research group: Daniele Yang-Yao, Marco Sazzini and Alessio Boattini.

Thanks to Francesca Brisighelli, George Busby and Cristian Capelli of the University of Oxford for the support, intellectual and financial, in the genotyping of the samples and the advices for the analysis.

Thanks to Prof. Giovanni Destro-Bisol for his support, the patience showed over the years and more importantly his critics and his precious advices for this thesis. Thanks to Valentina Coia, without her there wouldn't be any project regarding Greece. Thanks to all the other people of the laboratory of Molecular Anthropology of University "La Sapienza" of Rome: Francesco Montinaro, Veronica Marcari, Marco Capocasa and Valeria Montano.

Thanks to Doctor Maria Koukoura of the Hospital of Chalkida and Doctor Kallistheni Farmaki of the Hospital of Korinth and their staffs for their important help and support during the sampling.

Finally, a great thanks to all the donors that kindly offered their DNA for this research