Mitochondrial DNA diversity and origin of human communities from 4th- 11th century Britain.

Topf, Ana Laura

Mitochondrial DNA diversity and origin of human communities from 4th - 11th century Britain.

A thesis submitted to the University of Durham

for the degree of Doctor of Philosophy

2003

Ana Laura Töpf

School of Biological and Biomedical Sciences

University of Durham

Supervisor: Dr. Rus Hoelzel

Abstract

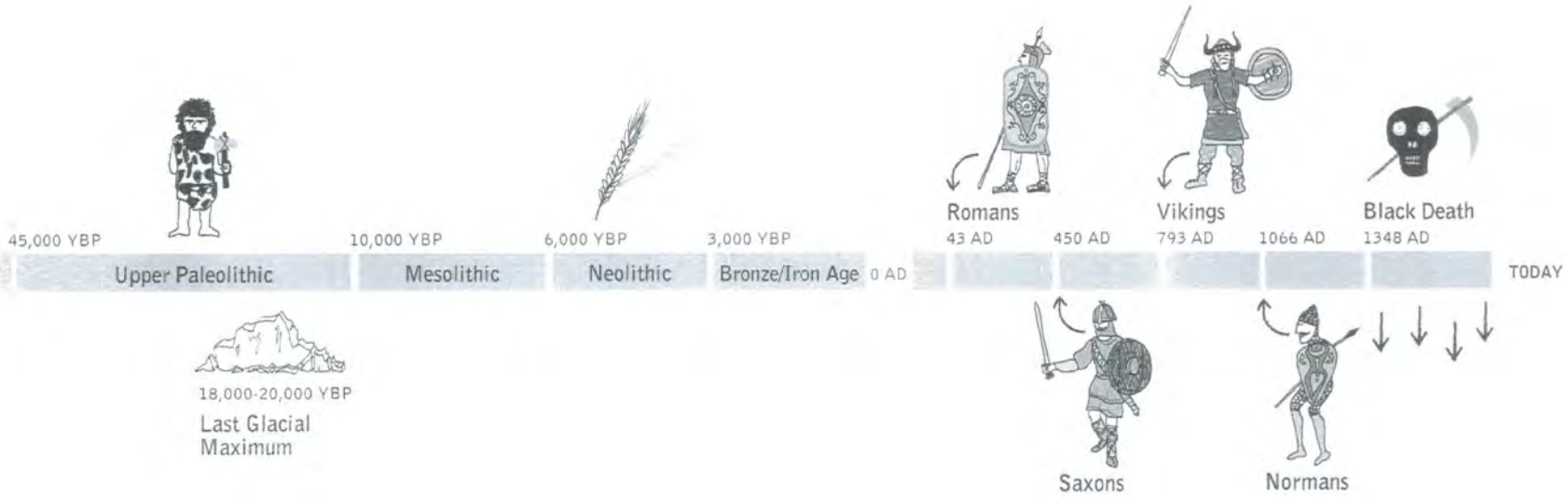
Neither the archaeological nor the historical data have yet allowed a full understanding of the nature of the Germanic settlement in England. Analysis of the genetic structure of past history has mostly been carried out by inference from extant populations. However, genetic flow through migration over time is likely to have altered the genetic composition of modern samples. Analysis of the genetic composition of ancient populations (provided the authenticity of their DNA is obtained) gives a direct sight into the past. Thus, mitochondrial DNA from pre-Saxon (4th century), early Saxon (5th-7th century) and late Saxon (9th-11th century) settlements has been analysed to obtain a better understanding of the population history of Britain.

A methodology has been optimised, by which, ancient DNA from 1,000-1,800 year old archaeological material was extracted and ~200-bp fragments of the HVS-I, amplified and sequenced. Rigorous controls for work in human ancient DNA were undertaken to prevent and recognise contamination. Established authenticity criteria were followed, including expected ancient DNA behaviour, internal replication of sequences and confirmation by independent labs. The sample size obtained has enabled a population-level study of communities of ancient Britain. In addition, an extensive database of >6500 mitochondrial DNA sequences was compiled for comparisons.

Several estimates of haplotype and nucleotide genetic diversity were computed for modern and ancient populations. Counter-intuitively, the modern population of England, encompassing all successive waves of migration to the island, has a lower diversity than the ancient population, suggesting that diversity has been lost over the last millennium. In addition, mtDNA genetic continuity between ancient and modern England seems to have been interrupted. Founder analyses of early (5th-7th century) and late (9th-11th century) periods indicate that, whereas the late period seems to have had Viking genetic influences, the early period has no close relationship with Germanic populations. Instead, the females of the early Anglo-Saxon period seem to represent the native British population. The female contribution of the Anglo-Saxon invasion would have therefore been minor, at least at that time and at these sites.

The close genetic affinity between the ancient British population and the northernmost populations of Europe suggests they might have shared a common past during pre-history. It is proposed that, after post-glacial times, inhabitants of areas now submerged expanded to northern territories. The early settlements analysed reflect that very early expansion. Some time since then, reduction in diversity seem to have occurred (possibly due to variation in family size after repeated epidemics) leading to the present day mtDNA composition of England.

Timeline of Britain



To all who ask themselves
who we are,
where we come from
and where we are going to

Table of Contents

List of Figures

Chapter I – Introduction

Chapter II – Materials and methods

Chapter III – Optimisation of methods to retrieve poorly preserved DNA

Chapter IV – Genetic diversity of mtDNA from ancient England

Chapter V – Geographical origin of mtDNA from ancient England

List of Tables

Chapter I – Introduction

Chapter II – Materials and methods

Chapter III – Optimisation of methods to retrieve poorly preserved DNA

Chapter IV – Genetic diversity of mtDNA from ancient England

Chapter V – Geographical origin of mtDNA from ancient England

Declaration

No part of the work contained within this thesis has been previously submitted for any other degree or qualification at the University of Durham or any other university or institute of learning. All the work was performed by the author, except where stated otherwise.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. A. Rus Hoelzel, for very helpful advice and discussion during this research and patience to read my drafts. I would also like to thank my former supervisor Dr. Charlie H. Shaw for giving me the opportunity to re-start my PhD in the Department of Biological Sciences and for his support regarding the forensic work. I feel I also must thank my initial supervisor Dr. Malia Fullerton, who gave me the opportunity to come to Durham in the first place, and who taught me quite a lot in a rather short time, and finally the EAAF (Equipo Argentino de Antropología Forense) for having trusted in me for the Genetic Identification Project.

I am grateful to the Committee of Vice-Chancellors and Principals of the UK for an Overseas Research Student Award which covered most of my tuition fees and to the Fundación Antorchas (Argentina) and The British Council for a partial scholarship during these years. I would also like to thank Dr. Alan Bilsborough (Department of Anthropology, Durham) and The Ian Karten Charitable Trust for financial help during the first and second and third year of my PhD, respectively, and Dr. Richardson and The Graduate Society for an Accommodation Award.

I owe debt to Trevor Anderson (Canterbury Archaeological Trust), John Lucas (Jewry Museum), Mandy Marlow (Department of Anthropology, Durham), Lorraine Mephan (Wessex Archaeology) and Elizabeth Popescu (Norfolk Archaeology Unit) for giving me access to the archaeological material used in this investigation and specially to Dr. Rob Fleischer and Tom Gilbert for replicating independently a number of samples, which greatly contributed to the validation of this research.

I would also like to thank

- Dr. Sussanne Hummel (Universität Göttingen, Germany), Dr. Martin Evison (University of Sheffield), Dr. Donga Yang (McMaster University, Canada), Dr. Mark Thomas (University College London), Dr. Chris Dudar (McMaster University, Canada) and specially to Dr. Michelle Harvey (University of Washington, USA) and Dr. Rafael Montiel (Universidad de Barcelona, Spain) for their help regarding methodological aspects of the aDNA work.
- Dr. Vincent Macaulay (Oxford University), Dr. Agnar Helgason (deCode Genetics, Iceland), Dr. Francesc Calafell (Universidad de Barcelona, Spain) and Dr. Peter Forster

(Cambridge University) for access to their mtDNA data and methods and for help on the use of the program NETWORKS 2.0;

Dr. Ben Dodds (Department of History, Durham) for advice regarding the Black Death, Dr. Malcolm Smith (Department of Anthropology, Durham), Dr. Sam Lucy (Department of Archaeology, Durham) and especially Dr. Heinrich Härke (University of Reading) and Dr. Caroline Wickham-Jones (University of Newcastle) for their invaluable help regarding archaeological aspects of this research;

Gillian Storey and Vicky Kelly (DNA sequencing lab) for their patience with my ancient DNA samples, Paul Sidney for the pictures and Manuel Weber, Dr. Anna Fabiani and my brother Leo for statistical advice.

and specially Alma Larroca for the timeline, Dr. John Hamilton for proof-reading, Emma Kestelboim for comments and José Töpf for edition.

I would also like to thank Anna Fabiani, Demetrios Vassilakos, Stefania Gaspari, Ada Natoli, Courtney Nichols, Colin Nicholson, Dan Engelhaupt and Carlos de Luna López, who made my long data-analysis days in the lab more enjoyable. Outside the lab I shared good times with Maria Angelica and Claudia, and I want to thank them for their young and already very close friendship.

Finally, I want to thank Mr. Kevin Daly for his help during these years and Carina for being always there for me. I am in debt with Nora for her invaluable support during difficult moments.

I feel I must also mention my friends Alma, Valeria, Alina and Irene for being with me during all these years – miles and oceans will never change our friendship.

I owe John many many thanks for his extremely helpful advice and discussions, and above all, for his loving support and immense (but not unlimited!) patience during this stressful time. Thanks for those moments when I could actually forget about work.

Lastly, I have no words –in English or Spanish– to express my gratitude to my parents for their unconditional support during all these years. Their encouragement and love kept me going during the most difficult times. I deeply thank them.

Abbreviations

A	adenine
AD	Anno Domino
AMOVA	Analysis of Molecular Variance
aDNA	ancient DNA
bp	base pairs
BC	Before Christ
BP	before present
BSA	bovine serum albumine
ca.	circa
C	cytosine
C.	century
dNTPs	2'- deoxyribonucleoside 5'- triphosphate
DNA	deoxyribonucleic acid
DTT	dithithreitol
EDT	ethylenediamine tetracetic acid
EtBr	ethidium bromide
g	aceleration due to gravity
G	guanine
GuSCN	guanosine thyocianate
H	heterocygosity
hg	haplogroup
ht	haplotype
kb	kilobases
mtDNA	mitochondrial DNA
PCR	polymerase chain reaction
T	thymine
Tm	melting temperature
Tris	2-amino, 2-(hydroxymethyl)propane 1,3-diol
U	units of enzyme activity
UV	ultraviolet
v/v	volume to volume
V	volts
W	watts
w/v	weight to volume
YBP	years before present
KY	Kilo years (i.e. 1,000 years)

Chapter I - Introduction

The story of early Britain has traditionally been told in terms of waves of invaders. Being either Stone Age or Neolithic people, Saxons, Vikings or Normans, they would displace their predecessors to found a new civilization. For over tens of thousands of years people have been moving into –and out of– Britain, sometimes in substantial numbers. Nevertheless, whether or not there has been a basic continuity of population still remains unclear.

Dichotomies such as acculturation vs. extermination, indigenism vs. colonization, regionalism vs. replacement seem to be a common motif, not only in British, but also in European history in general. However, this is very likely to be a simplification, with the human story being much more intricate. The puzzle is how to unravel that past.

Events of human history have mostly been addressed through interpretations of archaeological, paleontological, anthropological and speleological evidence. More recently, molecular genetic techniques have also been applied to understand the human past, opening up a new field (Renfrew, 2000). Much valuable information has accumulated from the analyses of DNA from living populations (e.g. Cann *et al.*, 1987). However, contemporary DNA provides only indirect evidence of the historical processes which have lead to present-day genetic composition of species and populations. Analyses of DNA from ancient populations, on the other hand, give a truly direct insight into the past.

Understanding of the British first millennia has mainly come from the study of historical, archaeological and linguistic data. Written sources (e.g. Bede, 731 AD; Gildas, ca. 550 AD), grave-good contents (e.g. Lucy, 2000; Härke, 2001) and place-names (e.g. Wainwright, 1962) have provided most of what we know today about the people living in Britain at that time. Inferences from genetic analyses have also greatly contributed to this knowledge (Wilson *et al.*, 2001; Weale *et al.*, 2002) yet direct investigation of ancient British communities has never been successfully undertaken before.

I.1 Human migrations in Europe

The pattern of genetic diversity currently observed in modern populations in Europe has been interpreted to reflect three large-scale phenomena recorded in the archaeological record: the early expansion of anatomically modern humans (AMH) during the early Upper Paleolithic, the post-glacial re-colonization(s) of Europe after the peak of the Last Glacial Maximum and the posterior spread of Neolithic farmers from the Near East. This Chapter will give acquaintance with those early events and discuss the different views regarding their importance in the modelling of the genetic pool of Europe today. The application of mitochondrial DNA (mtDNA) and, in particular of ancient DNA (aDNA), to the understanding of this issue will be also discussed.

I.1.1 Geological and archaeological definitions

The story of the genus *Homo* begins in the Pleistocene around 2.5 million years ago but the divergence of modern humans *among themselves* may be only 100,000 years-old (Stringer, 2002). To put this in temporal perspective, Table I.1 shows the geological subdivisions of the last 1.5 million years and the corresponding development of humans. Standard geological definitions include the Pleistocene and the Holocene Epochs. The Pleistocene can be divided into a number of stages marked by climate change. These correspond to cold and warm phases referred to as glacials and interglacials, respectively. In Britain, the cold stages are known as the Anglian, Wolstonian and Devensian, with the Hoxnian and Ipswichian warm stages between them. During the Devensian, i.e the last glacial stage, the point at which the global ice extent was at its greatest, approximately 18,000-20,000 years ago, is known as the Last Glacial Maximum (LGM). The LGM was much more arid than present, almost everywhere, with desert and semi-desert occupying huge areas of the continents and forests shrunk back into refugia. After this there was a rapid global warming, marking the beginning of the Flandrian warm period in which we live today (Dansgaard *et al.*, 1993).

Era	Starting date	Development of humans
Pleistocene		
Early	1.7 mya	*Homo habilis – Homo erectus*
Middle	700 kya	*Homo erectus – Homo sapiens*
Late	130	Anatomically modern humans (AMH) in Africa and Asia
'Later'	50 kya	AMH in Europe (after Neanderthals) Australia and America
Holocene	10 kya	Beginnings of agriculture

Table I.1 Geological subdivisions and the corresponding development of humans. (mya) and (kya) are million and thousand years ago, respectively (taken from Cavalli-Sforza *et al.*, 1994).

Archaeological definitions include the Paleolithic and Neolithic, which correspond only approximately to Pleistocene and Holocene designations. The Paleolithic is subdivided into Lower, Middle and Upper periods –and these are similar but not identical to those of the Pleistocene. This is because the archaeological classification is based on a different criterion: the type and degree of development of stone industries. The sequence and periods of stone industries vary from place to place, as do the definitions of the Paleolithic and its periods. To avoid confusion, Clark (in Tatersall *et al.*, 1988) proposed technological models to describe the changes in stone industries associated with the different periods (see Table I.2).

The Upper Paleolithic, starting in Europe ~50,000 YBP, is further divided into Early Upper Paleolithic (EUP), Middle Upper Paleolithic (MUP) and Late Upper Paleolithic (LUP). The transition between the LUP and the Neolithic is known as Mesolithic, a term that has been sometimes loosely used –mostly by non-archaeologists– to define the post-glacial period (e.g. Richards *et al.*, 2000; Simoni *et al.*, 2000a; Barbujani & Bertorelle, 2001). As abovementioned, Paleolithic and Neolithic are not contemporaneous in all parts of Europe, so their dating is relative. For clarity, a time-scale modified from the one proposed by Richards et al. (2000) based on archaeological and paleo-climatological information (Tatersall *et al.*, 1988; Dansgaard *et al.*, 1993; Strauss, 1995) has been adopted for the relevant periods mentioned in this study.

Period	Date	Culture	Stone Industry
Paleolithic			
Lower	>1.7 mya	Oldowan - Acheulan	1. simple flakes - 2. formally shaped tools
Middle	200 kya	Mousterian	3. flakes struck with prepared cores
Upper			4. blades and burins
Early (EUP)	50 kya	Aurignacian	
Middle (MUP)	25 kya		
Late (LUP)	18 kya	Solutrean[a] (*LGM*)	5. microliths and composite tools
Mesolithic	15 kya	Maglemosian[b]	
Neolithic	10 kya		6. ground-stone tools

Table I.2 Archaeological periods and some of the major associated cultures over the last 1.7 million years. Dates are given as approximate 'starting dates' in Europe. (mya) and (kya) million and kilo years ago, respectively. (a) mostly found in France, (b) mostly found in Scandinavia. Stone industries 1-6 correspond to those proposed by Clark (1988). The last Glacial Maximum (LGM) is indicated.

I.1.2 Origins of modern humans – the Paleolithic

The beginning of the Upper Paleolithic 40,000-50,000 years ago marks the first appearance of anatomically modern humans (AMH) in Europe (Mellars, 1989; 1993), prior to which it had been occupied by the Neanderthals for about 250,000 years (Stringer, 1993) and other archaic humans for nearly one million years (Carbonell *et al.*, 1995; Gutin, 1995; Pares & Pérez-González, 1995). Cultural change was dramatic and it is referred to as the transition, or even revolution, from the Middle Paleolithic culture (the Mousterian) to the Upper Paleolithic culture, whose fully-fledged cultural tradition in Europe and the Near East was the Aurignacian (Mellars, 1989; 1992; 1993) (see Table I.2). Figure I.1 shows the radiocarbon evidence for the geographical distribution of early Upper Paleolithic sites (i.e. the Aurignacian industries) in Europe. This pattern was interpreted by Otte and Keeley (1990) (and incorporated in his argument by Richards *et al.*, 1997) to be consistent with a spread from east to west and a rapid replacement of the Neanderthal population (although Clark (1991) disagrees on this, based on the early dates of the Spanish sites; see Fig. I.1).

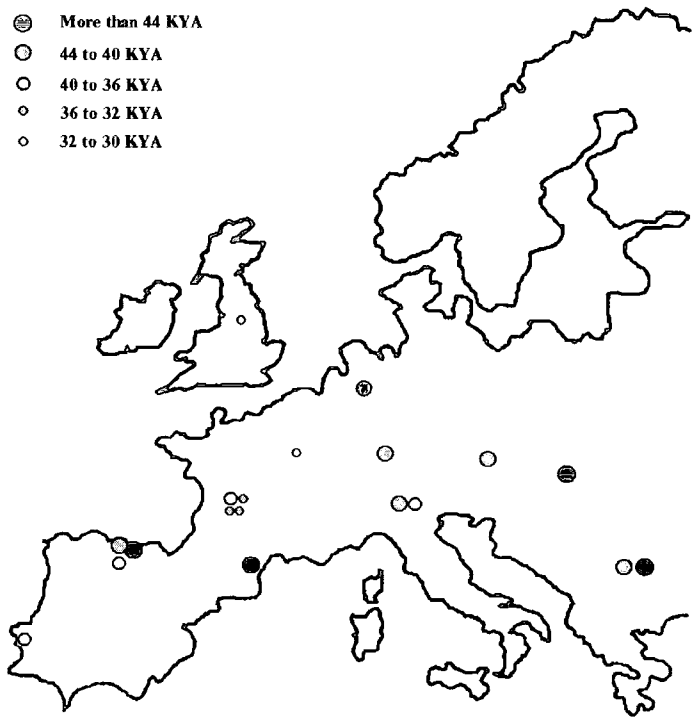


Figure I.1 Radiocarbon evidence for the spread of the early Upper Palaeolithic in Europe (taken from Richards *et al.*, 1997).

The nature of the transition between Neanderthals and AMH has been the issue of great debate, with the two main theories known as the 'Multiregional' and the 'Out of Africa' models. The multiregional maintains that our ancestor, *Homo erectus,* moved out from Africa >1 million years ago and that *Homo sapiens* evolved gradually from *Homo erectus* worldwide by the effects of genetic drift and natural selection (Thorne & Wolpoff, 1981; Wolpoff *et al.*, 1984). The other theory, the out-of-Africa theory proposes that *Homo sapiens* originated in Africa 100,000-200,000 YBP and that all present human populations outside Africa are descendants of a population that moved from Africa and replaced the archaic humans (Rightmire, 1984; Stringer *et al.*, 1984; Stringer & Andrews, 1988).

In the multiregional theory, current human genetic variation is seen as the product of the early radiation of *Homo erectus* from Africa. The expectations under this model include a pattern of continuity –genetic and morphological– from these middle Pleistocene hominoids to present day humans, high interpopulation differences (greater between peripheral areas), and intrapopulation variation greatest at the centre of human

range. On the other hand, the out-of-Africa theory would predict continuity *only* since the appearance of *Homo sapiens*, relatively low interpopulation differences (greater between African and non-African populations), and intrapopulation variation greatest among African populations (Stringer & Andrews, 1988).

The out-of-Africa theory was strongly supported by a phylogenetic analysis of restriction fragments length polymorphisms (RFLPs) data of mitochondrial DNAs (mtDNAs) sampled from different parts of the world (Cann *et al.*, 1987). The phylogenetic tree obtained had two primary branches; one of them was composed entirely of Africans and the other included all the populations studied (see Fig. I.2).

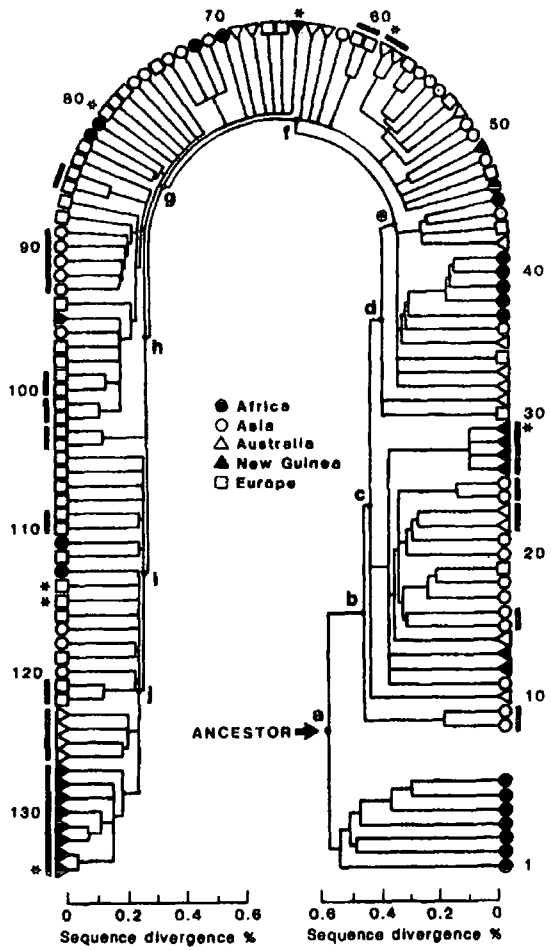


Figure I.2 Phylogenetic tree for 134 types of mtDNA from several continents, based on RFLP data (taken from Cann *et al.*, 1987).

Based on this, the authors inferred that Africa was a likely source of the human mtDNA gene pool. The fact that Africans proved to be by far the most diverse group among the peoples of various continents also supported an African origin for the AMH, as predicted by the out-of-Africa model (Stringer & Andrews, 1988). Previous mtDNA studies (Johnson *et al.*, 1983) also found that the Africans possessed the greatest variability among the populations studied. In addition, studies based on nuclear classical markers, such as blood groups and serum proteins (Nei & Roychoudhury, 1988) and RFLPs of nuclear DNA (Wainscoat *et al.*, 1986) also suggested an African origin for the human nuclear gene pool. Assuming that mtDNA divergence accumulates at a constant rate in humans (and estimating it as ~2-4% per million years) Cann et al. (1987) proposed that contemporary human mtDNA traced back through maternal lineages to an ancestral mtDNA present in an African population some 200,000 years ago –the well-known mitochondrial Eve. This final conclusion is hard to reconcile with the multiregional model.

However, Cann et al. (1987) attracted criticism; among the perceived weaknesses were that a) an indirect method was used to compare DNA, namely RFLPs, b) the same data could have been used to draw up many different trees (i.e. the tree had no statistical confidence), c) the strategy to root the tree was inadequate and d) the estimation of the mutation rate was inaccurate. Vigilant et al. (1989; 1991a; 1991b) addressed these issues by direct sequencing of ~600 bp of the control region of the mtDNA of 189 individuals. The data were used to construct a phylogenetic tree rooted by the outgroup method (which is preferable because it does not rely on the assumption that the evolution rate is equal for all the lineages) and the statistical support was evaluated by two methods (both with highly significant values). To better estimate the rate of evolution, *Pan paniscus* was used as outgroup and the divergence between chimp and human mtDNA sequences calibrated against the divergence time (estimated in 4-9 million years ago; Hasegawa *et al.*, 1990). The obtained divergence rate, 11.5% per million years was used to calculate the time to the most recent common ancestor (MRCA). None of these results contradicted the original data and were therefore consistent with an African origin of AMH within the last 200,000 years, with subsequent migrations out of Africa that established human populations in Eurasia.

Most of the controversy over the implications of the results by Cann et al. (1987) and Vigilant et al. (1991a; 1991b) stems from the dating point. If the ancestor was much older (in the order of 1 million years ago) the migrations out of Africa evident from the mtDNA phylogeny would correspond to the well-known movement of *Homo erectus* from Africa into Eurasia. The absence of mtDNA sequences that diverged from one another more than ~280,000 years ago led to the conclusion that the migrating population probably replaced the resident Eurasian populations that descended from earlier migrations of *Homo erectus* from Africa 800,000-1,000,000 YBP (Vigilant *et al.*, 1991a).

In addition, ancient DNA analyses from Neanderthal specimens have been considered to support the out-of-Africa model (Lindahl, 1997; Ward & Stringer, 1997), as there seems to be no continuity between the genetic pool of those archaic hominids and living humans. So far, mtDNA sequences from three Neanderthal specimens have been obtained: the original specimen from the Feldhofer Cave in Germany (Krings *et al.*, 1997; 1999), a specimen from the Mezmaiskaya Cave in northern Caucasus (Ovchinnikov *et al.*, 2000) and one from the Vindija Cave in Croatia (Krings *et al.*, 2000). These sequences are noticeably different from those of living humans, presenting overall >30 nucleotide differences when compared with the modern reference sequence (Anderson *et al.*, 1981). The average number of differences between any Neanderthal and modern sequence exceeds the amount found within living humans (Krings *et al.*, 2000). In a phylogenetic analysis, the Neanderthal sequences form a distinct group, belonging to a lineage that diverged earlier than the most recent common ancestor (MRCA) of living humans (see Fig. I.3). The results suggest that they have not contributed to the genetic pool of modern humans and that, although coexisting in Europe, Neanderthals and AMH would not have interbred to a significant extent (Cooper *et al.*, 2001). Stringer (see Kahn & Gibbons, 1997) argued that if the multiregional continuity was correct Neanderthal sequences should be closest to the Europeans, but results showed that Neanderthals were equidistant to all races.

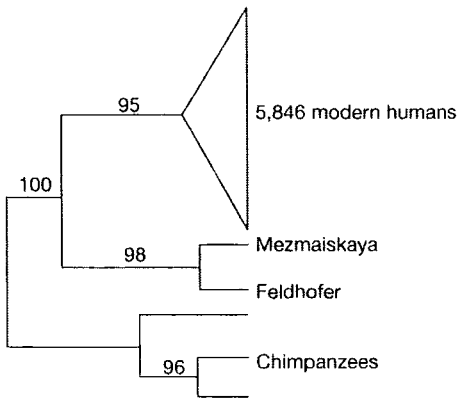


Figure I.3 Phylogenetic tree based on mtDNA data for Neanderthals, modern humans and chimpanzee (taken from Ovchinnikov *et al.*, 2000).

However, Relethford (2001) argues that the multiregional model does not predict that Neanderthal sequences will be more similar to Europeans than to any other region but instead that there will be no significant regional affinities. Absence of Neanderthal-like sequences in the modern gene pool might not be surprising given the high probability of extinction over tens of thousand of years (Norborg, 1998; Wall, 2000; Relethford, 2001). DNA analyses from a 60,000 year-old –but undoubtedly anatomically modern human– aboriginal Australian also cast doubt on the simple interpretation that the Neanderthal sequences support the out-of-Africa model (Adcock *et al.*, 2001). The mtDNA sequence obtained was also shown to fall outside the variation of modern humans, suggesting that both archaic and historical anatomically modern humans might have had mtDNA sequences different from that of humans today.

Although the interpretation of the Neanderthal mtDNAs can be controversial, it is generally understood to be most consistent with the replacement view of the out-of-Africa model. In addition, a recent extensive morphological analysis of crania from modern humans, Neanderthals and other Paleolithic specimens (i.e. early AMH) also supports the replacement model. Neanderthals showed a large morphological distance from and no strong morphological similarities to modern humans and AMH specimens (Harvati, 2003) suggesting that Neanderthals represent a different species (rather than a subspecies of *Homo sapiens*) which had no contribution to the evolution of modern humans in Europe.

If there was no Neanderthal contribution to the current European gene pool, European gene diversity should reflect the effects of demographic phenomena occurring since *Homo sapiens sapiens* arrived on the continent. By this reasoning, the first large-scale migration was the colonization of Europe during the Paleolithic around 40,000-50,000 YBP. The next population expansion is thought to have occurred some 25,000 years later.

I.1.3 Post-glacial expansions – the Mesolithic

Around 25,000 YBP the freezing temperatures of the Last Glacial Maximum (LGM) began, culminating, in northern Europe, in a peak cold condition about 18-20,000 YBP (Soffer & Gamble, 1990). By this time the British and Scandinavian ice sheets expanded to their maximum extend, roughly from the North Sea to the Kara Sea, covering most of Ireland, Scotland, England, Scandinavia and the east European plain, and the Alps and Pyrenees were totally covered by glaciers (Soffer, 1990) (see Fig. I.4). The proximity of the ice sheets subjected northern Europe to extreme periglacial conditions, creating a habitat which has been considered too inhospitable for man (Strauss, 1990).

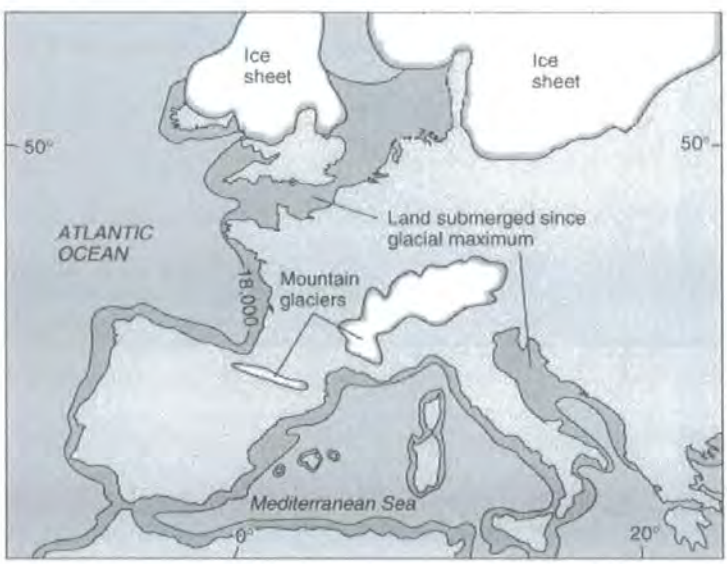


Figure I.4 Approximate extent of the ice sheets and glaciers at the peak of the Last Glacial Maximum (LGM) around 18,000 YBP. The limit of the Devensian and Weichselian (i.e. British and Scandinavian ice sheets) is controversial (see Chapter V) and is shown here only as a reference to the rest of Europe (redrawn from Frenzel *et al.*, 1992).

Archaeological evidence suggests that Britain and Belgium, as well as other northern parts of Europe, were apparently unoccupied between 25,000 and 14,000 YBP (Mellars, 1974; Evans, 1975; Otte, 1976). It is generally believed that during this period populations abandoned most parts of northern Europe and travelled south searching for a warmer climate and sheltering themselves in refugia (Jochim, 1983; Soffer, 1987; 1990; Kozlowski, 1990). Based on archaeological data, the two major refugia are thought to have been to the west, in south-western France and Cantabria (Strauss *et al.*, 1980; 1990; Jochim, 1983; 1987) and to the east, along major rivers of the Ukraine and the Central Russian Plain (Soffer, 1985; 1987) but other minor refugia could have also existed (Dolukhanov, 2000). The very scarce archaeological record during 17-20,000 YBP outside the refugia areas seems to support this idea (Housley *et al.*, 1997), yet it is impossible to categorically show a lack of human occupation in northern Europe. This paucity in archaeological finds might have been the result of a decline in overall population or, alternatively, the archaeological record did exist but has been removed by environmental factors, such as heavy erosion (Weniger, 1989).

Nevertheless, by ~15,000 YBP, when the ice started its retreat and climate became warmer, northern sites become frequent again, which has been commonly recognised as the 'repopulation' of northern Europe (Housley *et al.*, 1997). If so, reoccupation of Europe must have occurred as an expansion from the refugia where populations were sheltering. In fact, archaeological (Housley *et al.*, 1997) and genetic (Torroni *et al.*, 1998; 2001) evidence seems to indicate a post-glacial expansion from south-west France. Since the majority of the archaeological evidence comes from the western refugia, it has been believed that this refuge played the most important role in the repopulation of northern areas. However, as Soffer (1987) points out, the scarcity of archaeological finds in the east is mostly due to lack of archaeological work in that region rather than lack of archaeological evidence itself.

Post-glacial expansion into new territory has been previously suggested to be important in the geographic distribution of populations and genetic variants (Hewitt, 1989). Various genetic markers have been used to study the intraspecific phylogeographic patterns of several taxa (Taberlet *et al.*, 1998; Hewitt, 1999). Possible refugia and postglacial colonization routes have been thus identified for many species.

Although each species has its own colonization history, three broad patterns (or paradigm routes) of recolonization of northern Europe were observed: a) from a Balkan refuge, b) from a western/Spanish refuge, a central/Italian refuge and a eastern/Balkan refuge and c) from an Iberian and a Caucasian/Carpathian refuge (see Fig I.5).

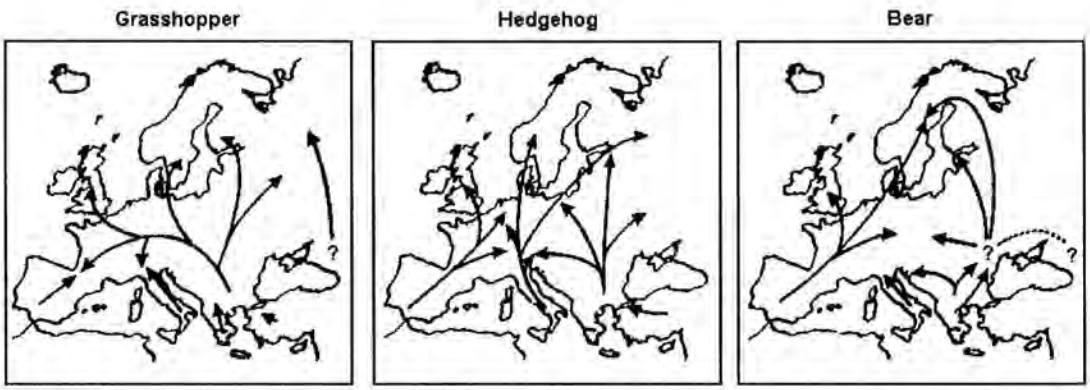


Figure I.5 Three paradigm postglacial colonizations from southern Europe deduced from DNA differences for the grasshopper, *Chorthippus paralellus*, the hedgehog, *Erinaceus europeus/concolor*, and the bear, *Ursos arctos*. The main refugial areas Iberia, Italy, the Balkans and the Caucasus contribute differently to the repopulation of northern parts (taken from Hewitt, 2000).

Britain received oaks, shrews, hedgehogs and bears from Spain but grasshopper, alder, beech and newts from the Balkans. Human reoccupation of Britain is less understood, although it has been traditionally believed that was the result of expansions from south-central Europe (Clark, 1932).

Italian genomes rarely populated northern Europe; it seems that the iced-capped Alps were an initial barrier to the northward expansion, while the Pyrenees apparently acted as a barrier in fewer cases. Conversely, despite being less studied than the western refuge, it appears that the Balkan/east Europe refuge have played an important role for most species in the recolonization of northern Europe (Hewitt, 2000). This may well have been the case for man too.

The term suture-zone has initially been introduced to describe 'the geographical overlap between pairs of species or sub-species that hybridize in the zone' (Remington, 1968); and the concept can be extended to the intraespecific level to characterise areas where the different populations met after a post-glacial expansion. Several suture-zones across Europe have been identified for a number species studied (Taberlet *et al.*, 1998; Hewitt, 1999). The four main suture-zones are shown in Figure I.6. The most obvious

ones corresponds to the Alpine barrier; the second suture-zone corresponds to the junction between populations of the Iberic refuge with populations of eastern origin (Balkans or other easternmost area) and is located somewhere between France and Germany (this zone is not localized exactly in the same place for different species); the third one is represented by the Pyrenees (and could be considered a particular case of the second) and the forth suture-zone is located in Scandinavia, which indicates that this area may have been colonized form the north and from the south for different populations originating from different refuges. So far, four mammalian species have been shown to support this latter suture-zone (Taberlet *et al.*, 1998). In addition, analyses of human allele frequencies also locate a zone of sharp genetic change –associated with linguistic boundaries– in approximate the same area (Barbujani & Sokal, 1990). Simoni et al. (2000b) using mtDNA data also recognised a genetic boundary which overlaps the one observed by Taberlet et al. (1998) (although they argued it might be due to the small size of the Norwegian sample).

Figure I.6 Suture zones defined by the main postglacial colonization routes in Europe (taken from Taberlet *et al.*, 1998).

Once Mesolithic communities settled back in northern Europe (<14,000 YBP), around 5,000 years later, another relevant phenomenon took place: the shift from a hunter-gather economy to the direct production of food.

I.1.4 Spread of Farming – the Neolithic

Agriculture arose in the river valleys in the Near East about 10,000 YBP (Clark, 1965a; 1965b; Perkins, 1969). Wild cereals were already a source of food, but farming allowed greater control of both quantity and quality of the yield and supplies. It is believed that components of early European farming including domesticated wheat, barley, sheep, goats and cattle were introduced into Europe from the Near East during the Neolithic (Ucko & Dimbleby, 1969; Thorpe, 1996). In support of this, genetic studies have indicated Near Eastern-origin of European cattle (Troy *et al.*, 2001). Archaeological evidence shows that the earliest sites of farming settlements were restricted to the Near East from where they expanded first, quickly spreading north of the Mediterranean through Greece and the Balkans, and then more slowly into the rest of Europe (see Fig.I.7) (Clark, 1965a; 1965b; Ammerman & Cavalli-Sforza, 1971). More recent data from radiocarbon dates agree with an expansion from the East but also show that the pattern was not equal across Europe (Pinhasi *et al.*, 2000).

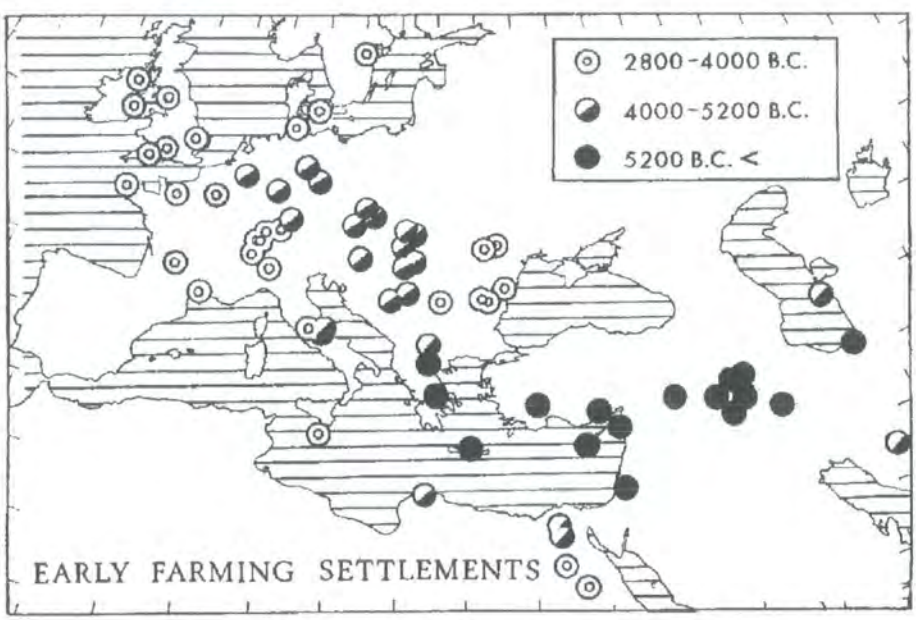


Figure I.7 Radiocarbon dates for the spread of farming. NOTE– Dates are quoted in years BC (before Christ) instead of years BP (before present) used throughout the thesis (taken from Clark, 1965b).

The routes and times of the spread of agriculture through Europe are relatively well established. However, its manner of spreading has been the subject of hot debate (Zvelebil & Zvelebil, 1988; Ammerman, 1989; Zvelebil, 1989). Two extreme

hypotheses have been proposed. The demic-diffusion hypothesis, proposed by Ammerman and Cavalli-Sforza (1971), postulates that the spread of agriculture was accompanied by extensive migration of Near Eastern farmers (referred to as the 'wave of advance'). In contrast, others have proposed a cultural diffusion model in which the transfer of agricultural economy occurred without significant population movement (referred to as a wave of assimilation) (Dennell, 1983; Barker, 1985; Whittle, 1996). There is clearly a spectrum of possibilities between these two hypotheses, with different contributions to the local indigenous development of farming by Mesolithic groups (Zvelebil, 1986).

The hypothesis of demic-diffusion is based on the idea that the spread of farming occurred because the local populations saturated the available resources. The first effect of agriculture would be the possibility of providing nourishment for far more people, which in turn led to growth in population density. Greater food availability allowed people to have many more children, providing more hands available to cultivate the land. On the other hand, hunter-gatherers would have an average of 5 children, one every 4 years. A 4-year gap means that the parents can always carry the youngest while the older ones are already able to walk and also, that children can be breastfed for longer, lowering the possibility of another pregnancy (Hayden, 1972). An average of five children per woman would keep the population stable, as conditions were harsh and more than half would generally die young (Cavalli-Sforza *et al.*, 1994). Under this view, Neolithic farmers in their spread would have outnumbered the Mesolithic Europeans. Ammerman and Cavalli-Sforza (1971) calculated the rate of spread of farming and the population growth as ~1km and ~3% *p.a.*, respectively.

However, Dennell (1983) argues that irrespective of the type of economy, subsistence groups are likely to have minimised risks by remaining within the areas they were familiar with. He also holds that hunter-gatherer groups were likely aware of farming communities elsewhere, and they were open to change customs. Under this view, the spread of farming over most of Europe (mainly the North) is explained by Mesolithic societies assimilating new techniques; an indigenous development of farming based on local and imported cultigens and domesticates, rather than expansion of farming pioneers (Dennell, 1983; Barker, 1985).

The genetic consequences of each model are rather different. Under the demic diffusion model, where the migrant farmers would have expanded with little admixture with Mesolithic communities, it is expected that most of the gene pool of the modern European population be derived from the farmers (Diamond, 1997; Barbujani *et al.*, 1998; Chikhi *et al.*, 1998b). On the other hand, under the cultural diffusion model, where the spread of farming would be based on trade and acculturation, the gene pool of prehistoric Europeans would have been essentially unmodified.

Genetic evidence in support of the demic-diffusion model initially came from analyses of classical gene-frequency data (Menozzi *et al.*, 1978; Cavalli-Sforza *et al.*, 1994). In a seminal paper, Cavalli-Sforza and colleagues constructed synthetic genetic maps of Europe and Near East, which are geographic maps of isopleths (i.e. lines of equal value) of principal-component values (PCs) calculated as optimised linear functions of all the gene frequencies (the original variable). To clarify this: a map for a single gene would be made of isogenic curves that link areas with the same gene frequency. In PC maps, all genes are considered together so that at each point, the values for individual genes are substituted by a single number, reflecting the whole group of genes under consideration. This number is the value of the synthetic variable at that point of the synthetic map. Hence, the method allows the analyses of large quantities of data and the identification of trends and patterns common to many genes that are the outcome of events influencing their geographic distribution. That is, a process of data reduction in order to explain most of the observed variance (Cavalli-Sforza & Cavalli-Sforza, 1995). Several PC values can be calculated by computing new synthetic variables (using the same set of data). Given than these variables are constructed so as to be statistically independent of one another, the isopleth map that each PC generates is independent of all other maps and explains a calculable fraction of the total genetic variation. Each map is ranked in relative importance by the fraction of variance that it is defined by its corresponding PC (Menozzi *et al.*, 1978).

It has been shown by simulation analyses that different PCs can separate expansions starting at different times and places (Rendine *et al.*, 1986). The rank of the PC is a function of the magnitude of genetic differences between the expanding populations. It also depends on demographic factors such as the rate of exchange with previous settlers

and the increase in population density. The rank of a PC is therefore associated with (but not exclusively dependent on) the antiquity of the expansion (Cavalli-Sforza *et al.*, 1993). It should be noted that population movements do not necessarily produce clines. To generate the observed gradient it is necessary that a) the expanding population differed genetically from the local population; b) the former were growing in numbers; c) they dispersed in a directional way, and d) they did not immediately incorporate other groups they met during the expansion into their communities (Barbujani & Bertorelle, 2001)

In their original paper, Menozzi et al. (1978) used gene frequencies for 38 protein loci to obtain the synthetic maps. Later on, the number was increased to 95 genes which yielded essentially the same results though with increased precision (Cavalli-Sforza *et al.*, 1993; 1994). Figure I.8 shows the first four principal component maps for Europe and the Near East, representing all together ~68% of the genetic variance.

The first PC map, accounting for 28.1% of the variation, showed the highest point centred in the Near East from where it diminished with distance towards the north-west. It was the striking topological similarity between this genetic map and the archaeological map of radiocarbon dates of the spread of farming from the Near East (see Fig. I.7 and Fig. I.8.a) which led to the formulation of the demic-diffusion model. That is, that the spread of farming from the Near East was a spread of farmers and not just technology.

Some of the early data of Menozzi et al. (1978) was also analysed by spatial autocorrelation; a method used to determine the dependence of the values of a particular variable on its values at geographically adjoining locations. A variable is correlated positively or negatively if its value at a given point in space is associated with its values at other locations. Spatial correlograms confirmed the existence of a pattern corresponding to that expected for a migration from the Near East to Europe (Sokal & Menozzi, 1982). An increased number of loci analysed with improved statistical techniques also corroborated this conclusion: Sokal et al. (1991) found that archaeological time distances (based on maps including dates for latest observed Mesolithic and earliest observed Neolithic settlements) and genetic distances were highly correlated.

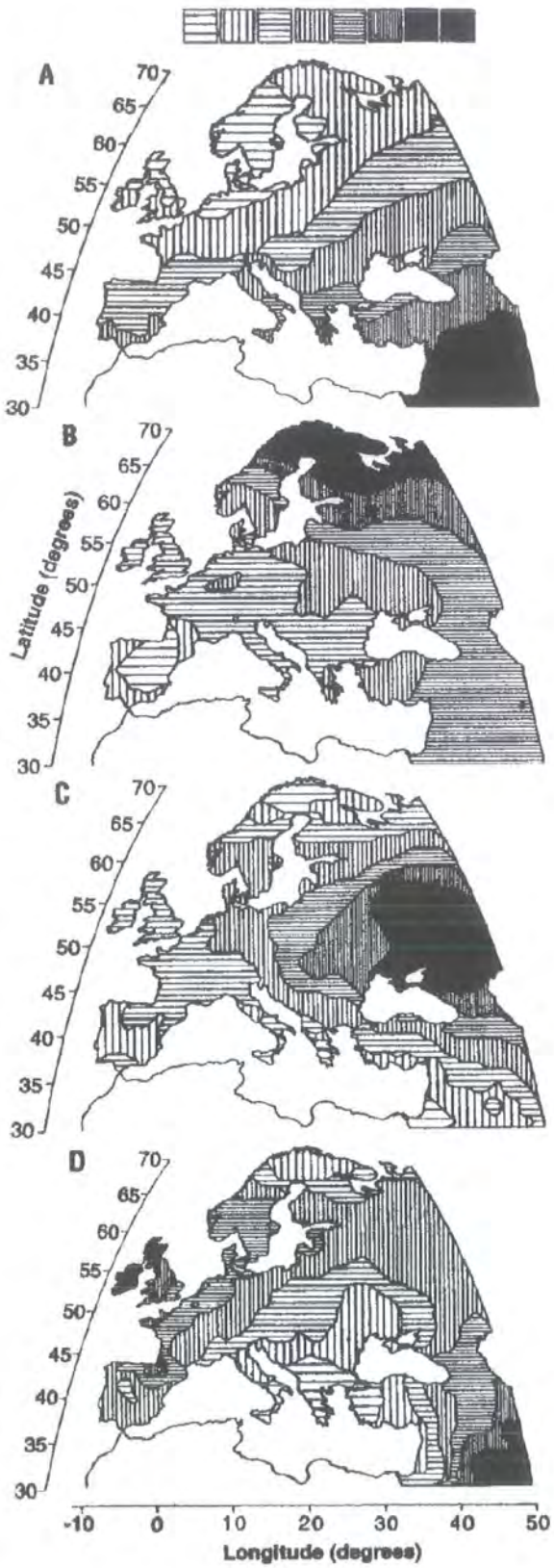


Figure I.8 Synthetic genetic maps of Europe. (**A** to **D**) The interpolated values of the first to fourth PCs, which represent 28.1, 22.2, 10.6 and 7.0% of the total genetic variation, respectively. The top legend represents the PC scores, ranked 1 to 8 on an arbitrary scale (taken from Cavalli-Sforza *et al.* 1993).

The 2nd PC map, accounting for 22.2% of the total genetic variation showed a concentric gradient like that of the first map but centred instead in the Iberian peninsula and expanding east-west (see Fig. I.8.b). The opposite pole showed a strong peak in the Saami population, which are not candidates for an expansion. Cavalli-Sforza et al. (1994) suggested several explanations for the observed cline, such as a climatic or ecological effect or an ethnic origin. Later, based on mtDNA analyses Torroni et al. (1998) proposed a post-glacial expansion from an Iberia refuge as a possible explanation for this cline (see later).

The original 3rd PC map (Menozzi *et al.*, 1978), was slightly modified when the analyses included more than twice the number of genes (Cavalli-Sforza *et al.*, 1993). The improved 3rd PC map, accounting for 10.6% of the total genetic variance, showed the highest peak in the Russian steppe at the north east of the Black Sea, from where it diminished concentrically (see Fig. I.8.c). This is the same general area where archaeological data place the beginnings of the Kurgan culture, believed to have started around 5,000 YBP by pastoral nomads. Kurgan is the Russian name for a burial mound, common in central Russian planes. Gimbutas (1970) suggested a connection between the Kurgan culture and migrations of speakers of Indo-European languages. The Russian steppe would have been the 'homeland' of Indo-European languages and several Kurgan expansions (between 4,300 and 3,000 YBP) would have been responsible for their spread into Europe. This hypothesis was adopted by Cavalli-Sforza and colleagues (1993; 1994) as explanation for the observed genetic cline in the 3rd PC map. Similarly, the expansion of the Scythians, who inhabited this area later on, was also imagined by the authors as another possible explanation (Cavalli-Sforza *et al.*, 1993; 1994). It is interesting to notice that the 3rd PC map might also account for a possible post-glacial expansion from the Caucasus area, as insinuated by Hewitt (2000) for other species (see section I.1.3).

The 4th PC map, accounting for 7% of the total genetic variance, is centred in Greece and its major colonies in the Aegea, western Turkey and southern Italy (see Fig. I.8.d). Note that in this map, unlike the previous ones, the centre of the expansion was considered to be located at the lowest rather than the highest frequency. The difficulty in interpreting this as a Greek expansion was that the Greek influence in classical times

never extended to the area covered by the second ring (of frequencies). Cavalli-Sforza *et al.* (1994) suggests that its 'propagule' in the eastern Balkans/western Ukraine might have actually been the place of the beginning of the expansion, which expanded to Greece and mixed with local inhabitants who later took the lead in the expansion and dominated the PC map. Again, the pattern observed would be in agreement with a post-glacial expansion from the Balkan area.

However, there is no reason to believe that any PC map in Europe is entirely due to one single expansion event. For example, as pointed out by Richards et al. (1997) the early Upper Paleolithic expansion of AMH in Europe also shows an east to west pattern (see Fig. I.1) and, as shown by Barbujani et al. (1995), computer simulations of expansion waves would yield similar results whether the expansion from the east occurred into a low populated Europe –as would be the case of a Neolithic expansion– or into an empty Europe –as would be the case of a Paleolithic first settlement. The absence of dating of the PC maps does not help to resolve this issue. It is hard to see why principal components and migrations and demic diffusion processes should correspond to each other in a one to one mode. Two or more migrations might contribute to a single component, and it may well be that, for example, the first component is a palimpsest of early Upper Paleolithic, Neolithic and perhaps other expansions from southwest Asia (Richards *et al.*, 1997). In this way, European genetic diversity should be seen as a result of several demographic processes that the populations went through.

Figure I.9 summarizes the main dispersals in Europe supposed to have occurred in the last ~40,000 years based on archaeological data: an early first colonization of AMH during the Upper Paleolithic, expansion(s) from refugia during post-glacial Mesolithic times (<18,000) and a later spread of farming during the Neolithic (~10,000). However, there is no consensus on the relative demographic impact of these migrations and on their importance in determining current patterns of geographic diversity of the genetic pool of Europeans. The controversy arose due to discrepancy between analyses based on allele frequencies of classical nuclear markers (e.g. Menozzi *et al.*, 1978) and those based on mtDNA sequences (e.g. Richards *et al.*, 1996).

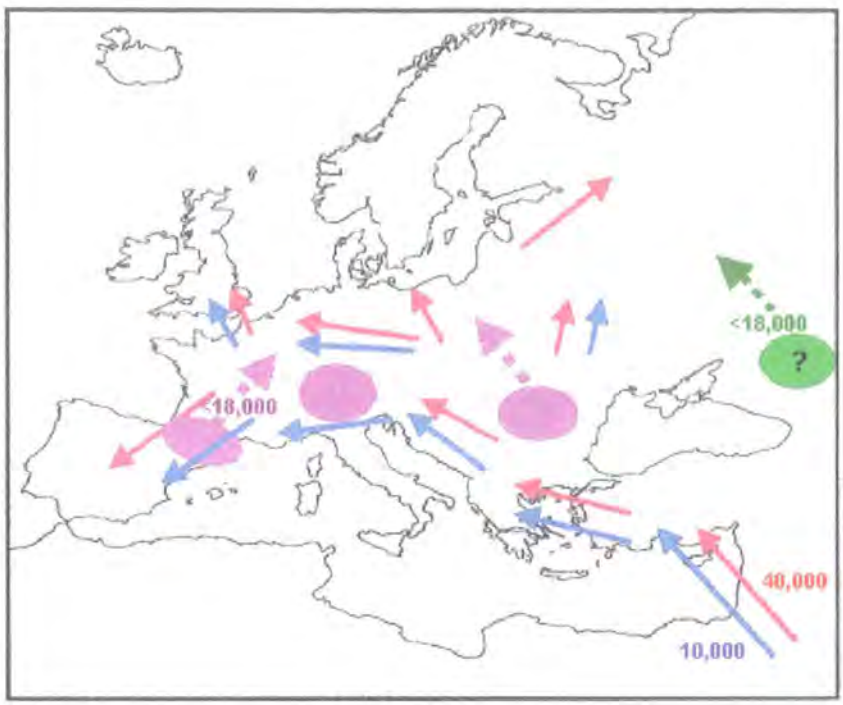


Figure I.9 Scheme of the main dispersal processes supposed to have occurred during the last ~40,000 years: the Paleolithic first colonization of Europe (*red arrows*) and the Neolithic demic diffusion (*blue arrows*). Approximate location of the western and centre glacial refuges are indicated (*violet ovals*), as are the possible Mesolithic expansion (*dashed violet arrows*) (the one from Iberia was proposed by Torroni *et al.* 1998). More speculative is the eastern glacial refuge, which is also represented (*green oval*) as is a possible expansion route (modified from Simoni *et al.*, 2000).

I.2 Mitochondrial DNA analyses

It has been argued by Wilson et al. (1985), that mtDNA is '*not just another molecule marker*' as it possesses the properties desired in an ideal molecular marker for phylogenetic analyses: a) a simple genetic structure, b) a straightforward matrilineal mode of inheritance without recombination, and c) a fast evolution rate (Avise et al., 1987). This combination of attributes can be used to provide perspectives on how, where and when the human gene pool arose.

Mitochondrial DNA studies have given insights into evolutionary processes such as our divergence with the apes (e.g. King & Wilson, 1975; Adachi & Hasegawa, 1995) and the origin of anatomically modern humans (e.g. Cann *et al.*, 1987; Vigilant *et al.*, 1991a; 1991b). Analysis of mtDNA has also been applied to study migration phenomena, such as the first colonization of the Americas and the Pacific (e.g. Forster *et*

al., 1996; Lum & Cann, 2000; Malhi & Smith, 2002, and Sykes *et al.*, 1995;Hagelberg *et al.*, 1999b) and the more recent peopling of Iceland (Helgason *et al.*, 2000). The impact of the re-colonization of South America by Europeans (Mesa *et al.*, 2000) and of the invasions in the North Atlantic (Helgason *et al.*, 2001) and the British Isles (Wilson *et al.*, 2001) was also investigated by mtDNA analyses. In addition, mtDNA has also been used to explore the genetic structure and affinities of ethnic groups such as Gypsies (e.g. Gresham *et al.*, 2000; 2001; Kalaydjieva *et al.*, 2001), Jews (Tikochinski *et al.*, 1991; Ritte *et al.*, 1993; Thomas *et al.*, 2002) and Maragatos (Larruga *et al.*, 2001).

I.2.1 Genomic organization

The human mitochondrial DNA is a circular molecule of ~16,500 base pairs, comprising 13 protein genes, 2 ribosomal RNAs and 22 transfer RNAs. The polypeptides encoded are all subunits of the mitochondrial energy-generating pathway, and together with the RNA genes encompass ~90% of the mtDNA genome. A large non-coding region spans approximately from nucleotide position 16024 to 576, known as the control region (CR). The remaining non-coding sectors are located between coding genes (Anderson *et al.*, 1981). Mitochondrial DNA evolves rapidly, with a mutation rate up to ten times higher than that of nuclear DNA (Vawter & Brown, 1986) giving rise to much variation between the mtDNA sequences from different individuals (Brown, 1980; Johnson *et al.*, 1983). The most variable region of the mitochondrial genome is the ~1,120-bp control region and the least variable is the coding 16S ribosomal RNA gene, less than a sixth as variable as the former (Aquadro & Greenberg, 1983; Cann *et al.*, 1984; Whittam *et al.*, 1986; Parsons *et al.*, 1997). Within the control region, the variation is concentrated in two regions: hypervariable segment I and II (HVS-I and HVS-II, respectively). HVS-I spans from 16024 to 16365, and HVS-II from 73 to 340, approximately (Stoneking *et al.*, 1991; Meyer *et al.*, 1999). Although useful for genetic identification analyses (Lutz *et al.*, 1996; 2000; Budowle *et al.*, 2002), the second hypervariable segment of the control region (HVS-II) offers little phylogenetic information, as there is extreme rate variation among sites in this segment (Aris-Brosou & Excoffier, 1996; Parsons *et al.*, 1997; Salas *et al.*, 2000; Malyarchuk & Derenko,

2001). The exception to this is site 73 which helps resolve the European phylogeny (Wilkinson-Herbots *et al.*, 1996) (see section I.3.1). In any case, most population studies have focused on the HVS-I (see Handt *et al.*, 1998; Röhl *et al.*, 2001).

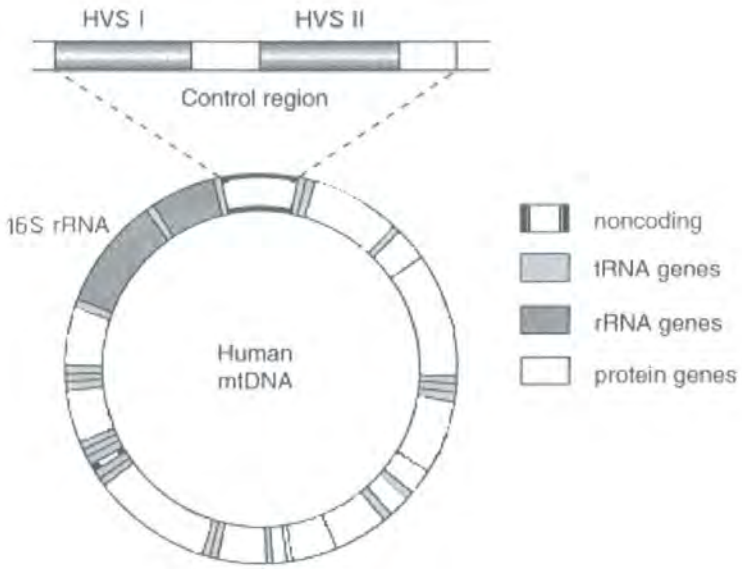


Figure I.10 Schematic diagram of the human mitochondrial DNA, showing the relative size and location of the control region and the hypervariable segments I and II (HVS-I and HVS-II). The least variable region, the 16S rRNA, is also indicated (taken from Hagelberg, 2003)

I.2.2 Mode of inheritance – recombination in mtDNA

Maternal inheritance of mtDNA in humans was first reported by Giles et al. (1980) from a family RFLP study, although no explanation of the process leading to this was suggested. It has been broadly accepted that, although sperm from vertebrates may contain hundreds of mitochondria in the mid-piece, this structure does not enter the egg during fertilization, and if paternal mitochondria do enter, very few, if any, survive the early cell division stages (Spuhler, 1988). Conversely, Ankel-Simons and Cummins (1996) pointed out that in the majority of mammals –including humans– the mid-piece mitochondria can actually be identified in the embryo, even though their ultimate fate is unknown. It has been proposed that the simplest explanation is that the few copies (~50-70) of paternal mtDNA (Ankel-Simons & Cummins, 1996) are diluted among the maternal copies (~100,000-200,000) (Reynier *et al.*, 2001) beyond recognition of the

techniques used, namely RFLPs (Hutchison *et al.*, 1974; Ankel-Simons & Cummins, 1996). However, a direct DNA sequencing study of father/child pairs did not show evidence for paternal transmission of mtDNA (Parsons *et al.*, 1997). It would seem that specific mechanisms have evolved to eliminate the mtDNA contribution of the spermatozoon in early embryonic development (Cummins, 2002) as has been reported for other mammals (Sutovsky *et al.*, 1999; 2000). In fact, leakage of paternal mtDNA has been detected in early stages of abnormal human embryos when dysfunctional oocytes were used for *in vitro* experiments (St John *et al.*, 2000).

Viable paternal inheritance of mtDNA has been observed in mussels (Zouros *et al.*, 1992), drosophila (Kondo *et al.*, 1992) and mice (Gyllensten *et al.*, 1991). In humans, although an early study detected paternal transmission of a mitochondrial pathology (Egger & Wilson, 1983), the inheritance of paternal mtDNA genome has been reported only recently (Schwartz & Vissing, 2002). Nevertheless, this single case does not prove that the paternal mtDNA can be stably inherited down the generations or that it can recombine with the maternal mtDNA.

On the other hand, the possibility of mtDNA recombination has been advocated by a number of authors, based on the high frequency of homoplasy (Eyre-Walker *et al.*, 1999a; Hagelberg *et al.*, 1999a; Morris & Lightowlers, 2000) and on the negative relation between linkage disequilibrium (LD) and distance observed between some RFLP sites in the mtDNA genome (Awadalla *et al.*, 1999; Eyre-Walker *et al.*, 1999b). However, some of these data either were shown to be incorrect (Macaulay *et al.*, 1999b; Hagelberg *et al.*, 2000) or could be rejected on a phylogenetic basis (Kivisild & Villems, 2000). Kivisild and Villems (2000) stressed that some of the RFLP sites which accounted for the short distance correlation reported in Awadalla et al. (1999) segregate in linkage with a number of other haplogroup-specific substitutions that are spread over the entire mtDNA genome (e.g. sites 7933 and 8391 of hg Y, see Torroni *et al.*, 1993; Schurr *et al.*, 1999). In addition, when other estimators of LD (Jorde & Bamshad, 2000; Kumar *et al.*, 2000) or more extensive data sets (Elson *et al.*, 2001) were used, no negative correlation between LD and distance was found, arguing against recombination in human mtDNA. It would seem therefore that there is no strong evidence to support the claim that mtDNA undergoes gametic recombination. However, the finding of the

abovementioned case of paternal inheritance of mtDNA (Schwartz & Vissing, 2002) has lead the advocates of recombination in mtDNA to re-open the debate; refering to the imagined paternal mtDNA ancestor as the 'mitochondrial Steve' (Bromham *et al.*, 2003).

I.2.3 Mutation rate – hot spots?

Two different approaches have been undertaken to estimate mutation and divergence rates (with the latter being twice the former). An indirect approach is to use the mtDNA sequences to construct a phylogenetic tree and then to date the branch lengths (i.e. the number of mutations that have occurred along the branch) by reference to an assumed molecular clock, calibrated with divergence dates estimated from the archaeological record (e.g. Hasegawa *et al.*, 1993; Tamura & Nei, 1993; Adachi & Hasegawa, 1995). In a more direct approach, pedigree analyses estimate mutation rates by simply counting the appearances of new mutations events in a number of generations. In this way, Howell et al. (1996) estimated a rate of 0.95 mutations/bp/Myr. This rate is ~8-fold higher than an average of previous phylogenetically derived rates, ~0.12 mutations/bp/Myr (e.g. 0.087 mutations/bp/Myr, Stoneking et al., 1992; 0.118 mutations/bp/Myr, Tamura & Nei, 1993; and 0.165 mutations/bp/Myr, Ward et al., 1991). Since Howell et al. (1996) used Leber hereditary optic neuropathy pedigrees (which primarily present mutations within the mitochondrial ND4 gene; see Howell & McCullough, 1990), it was questioned whether the disease state of the individuals might affect the mutation rate (Macaulay *et al.*, 1997; Jazin *et al.*, 1998). However, high divergence rates were also obtained when control region genealogies were used (Parsons *et al.*, 1997). Subsequent pedigree studies yielded lower mutation rates (Mumm *et al.*, 1997; Parsons & Holland, 1998; Sigurðardóttir *et al.*, 2000), yet they were always higher than the phylogenetically derived ones. It should be noted however, that the phylogenetically derived mutation rates were based on mtDNA differences between humans and chimps which depend heavily on assumptions such as the date of the deep ancestral divergence time. This may account for the low rates observed, as divergence over long periods of time is obscured due to mutational saturation (multiple mutational changes at a single site) (Kimura, 1968). In fact, phylogenetically derived rates based on

internal diversity within Amerindian populations yielded a higher value (0.18 mutation/bp/Myr) (Forster *et al.*, 1996). Nevertheless, several authors argued that the difference between 'phylogeny' and 'pedigree' rates was due to the presence of fast-mutating sites or mutational hot-spots in the CR (Pääbo, 1996; Macaulay *et al.*, 1997; Jazin *et al.*, 1998). Thus, pedigree analyses would predominantly detect those mutations which occur in a short period of time (i.e. hot-spots) leading to estimate high divergence rates. On the other hand, longer phylogenetic analyses would take into account both fast and slow mutation events over the time, which results in lower divergence rates. Pääbo (1996) suggested that neither the 'pedigree' rates nor the 'phylogeny' rates were right or wrong. Rather, they would be useful for estimating dates of different types: for comparisons of sequences that are closely related, sharing common ancestors on the order of hundreds or thousands of years, the pedigree rates would be better suited. For comparisons that go back hundreds of thousands or millions of years, the phylogeny rates would be better estimates. However, as with many other issues related to mtDNA, consensus has not yet emerged (Howell & Mackey, 1997; 2003; Cavelier *et al.*, 2000; Gurven, 2000). In addition, very recently, Hagelberg (2003) has suggested that hot-spots in the control region are actually ancient (and fixed) mutations, which are reshuffled among lineages by recombination in the mtDNA.

I.2.4 Levels of resolution and genealogy at the continental level

Before the appearance of complete sequences, the earliest work on human mtDNA began by digestion of the molecule, either with a single restriction enzyme in large numbers of samples (Denaro *et al.*, 1981) or with many enzymes in few samples (Brown, 1980). Subsequent studies tended to use 5 or 6 enzymes on fairly large sample sets (e.g. Johnson *et al.*, 1983). In this way, mtDNA studies could be broadly categorized on the bases of how much of the molecule was assayed. 'Low-resolution restriction analysis' constructed a restriction map of the molecule by cleaving with 5 or 6 restriction enzymes, whereas 'high-resolution restriction analysis' used 12 or 14. The results were usually presented as a phylogeography of global mtDNA variation. By high-resolution analysis, Cann et al. (1987) showed a deep split between African and

non-African populations, which, as mentioned and despite the criticism, was interpreted as evidence supporting the out-of-Africa model.

Early high-resolution RFLP studies applied to the analysis of one continent at a time showed that mtDNA could be classified into a small number of monophyletic clades defined by one or several restriction sites (Schurr *et al.*, 1990; Chen *et al.*, 1995). These clades or haplogroups (hg) are predominantly continent-specific. Thus, for example, in Africa, hg L (subdivided into L1, L2 and L3) encompasses >76% of mtDNAs (Chen *et al.*, 1995) and in Asia, the majority of mtDNAs are members of hgs A, B, C, D, G, E, Y and Z (belonging to super-hg M) (Torroni *et al.*, 1993). In the same manner, most Amerindian mtDNAs could be primarily classified into one of four (Asian) haplogroups, namely A, B, C and D (Schurr *et al.*, 1990), indicating that these haplogroups predated the colonization of Americas from Siberia (Torroni *et al.*, 1993). Later on, a fifth lineage, hg X, was also identified among native North American samples, representing 3% of the samples (Brown *et al.*, 1998). The origin and founder event of this Amerindian haplogroup has been under debate (see section I.4).

With regards to Europe, the first Caucasian markers were identified by Torroni et al. (1994), in a high-resolution RFLP analysis of Northern Americans of European origin (n = 175). Four haplogroups were determined, namely H, I, J and K, which seemed to be European-specific. However, associations between particular populations and haplogroups could not be carried out due to the diverse origin of the samples.

I.3 European mtDNA phylogeny

The first large-scale phylogeographic study of European mtDNAs was reported in a key paper by Richards et al. (1996). The authors directly analysed HVS-I sequences of >800 Europeans, belonging to 14 populations, and introduced the use of reduced median networks (RMNs) for their analyses. RMNs are specially useful to construct intra-specific phylogenies where sequences exhibit high homoplasy (i.e. parallel or recurrent mutation events), such as the case of human mtDNA (Bandelt *et al.*, 1995; 1999). In a RMN, haplotypes –either sampled or inferred as hypothetical intermediates– are represented by linked nodes (rather than terminal nodes of a tree) and the links between

them are associated with the evolutionary changes (i.e. mutations) which described the haplotype. As opposed to trees, RMNs display all the possible relationships among the individuals sampled (i.e. all the possible most parsimonious trees) and retain ambiguity due to parallel mutation.

The European network obtained portrayed a star-like phylogeny, in which the consensus sequence was by far the most frequent haplotype (at ~20%) (see Fig. I.11). This sequence happened to be identical to the Cambridge reference sequence (Anderson *et al.*, 1981). Most of the haplotypes were one or two mutations away from the core sequence, yet some branches protruded beyond this limit. Six major clusters or haplogroups were defined on the bases of specific segregating sites, some of them just a single mutation. The age of each cluster was estimated as a function of the internal diversity –calculated as the mean number of nucleotide differences (π)– and a divergence rate set to 1 transition per 10,500 years (for the entire HVS-I analysed; equivalent to a mutation rate of 0.17 mutation/bp/Myr) (note the previously mentioned discrepancies). The age of the clusters ranged between 6,000 and 50,000 years, with most of them dating back to the Paleolithic and only one cluster being more recently brought into Europe during the Neolithic (Richards *et al.*, 1996). This latter, cluster 2A (currently referred to as hg J) stood out as having lower diversity and clear Middle Eastern ancestry, and was therefore attributed to the Neolithic farmers. The frequency of this haplogroup in European populations was in average ~15% and hence the authors' conclusion that the current European mtDNA pool could be explained by a relatively small-scale Neolithic contribution and a much larger surviving Paleolithic component. This was the starting point of the Neolithic vs. Paleolithic debate (Cavalli-Sforza & Minch, 1997; Richards *et al.*, 1997).

As reviewed by one of the authors themselves (Sykes, 1999), some of the main objections to the interpretation of these results were that the phylogeny was not strongly supported and that the dating of the clusters was not appropriate.

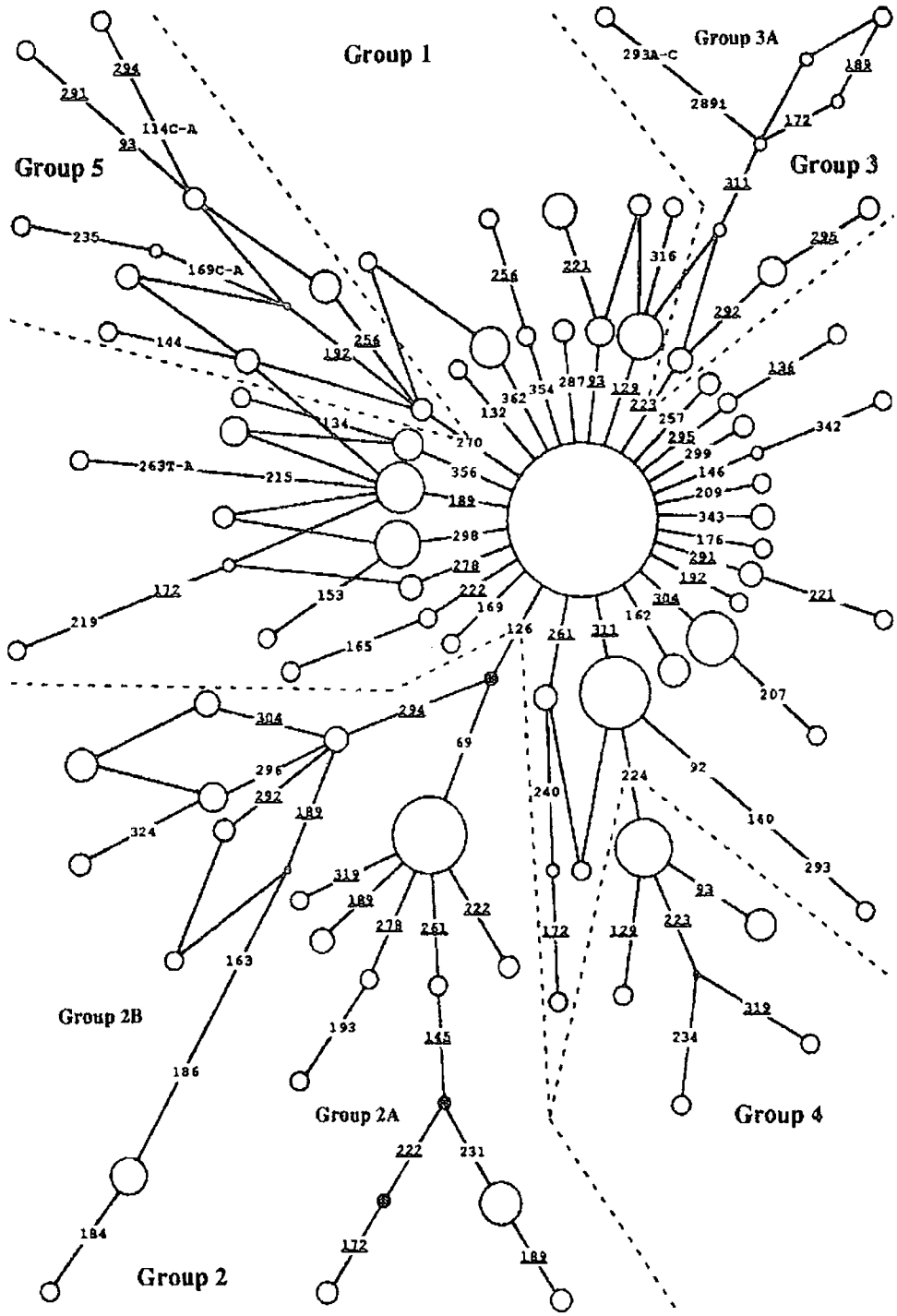


Figure I.11 Original European skeleton network, a reduced median network constructed from ~80 haplotypes occurring at a frequency ≥ 2 in Europe and Near East. Circles are haplotypes, with size proportional to their frequency. The filled small circles are haplotypes occurring only in the Near East. The numbers in the branches denote mutations, corresponding to the positions in the CRS (Anderson *et al.*, 1981) (minus 16000). All mutations are transitions unless specified (e.g. 243 A-T denotes an A→T transversion at position 243 and 289i denotes a cytosine insertion at position 290). Positions that are recurrent mutations are underlined. In any reticulation base notation is given only once but it is the same in parallel connections within it (taken from Richards *et al.*, 1996).

I.3.1 Improvement of the European phylogeny

The European phylogeny was improved by the incorporation of HVS-II sequences into the RMN construction. In particular, site 73 help resolve the phylogeny by splitting group 1 from groups 2 to 5. Therefore the long branches of clusters 2 to 5 did actually not protrude from the CRS but from a more ancestral type (Wilkinson-Herbots *et al.*, 1996). In addition, analysis of larger data sets comprising more geographical zones showed that group 1 could be further divided into two clusters and that several sub-clusters existed within group 3 (Richards *et al.*, 1998).

More information was obtained from the second high-resolution RFLP analysis of (mostly) coding regions of three European populations, namely Finns, Swedes and Tuscans (n = 134). The entire mtDNA genome was first amplified in 9 overlapping PCR fragments which were then digested (separately) with 14 restriction enzymes in order to generate respective RFLP profiles (no probes were used). Based on these composed profiles (characterised by absence/presence of restriction sites), another 5 haplogroups were identified, called T, U, V, W and X (Torroni *et al.*, 1996). A very low frequency of Asian hg M was also detected. It was thus suggested that 99% of the mtDNA in Europe could be subsumed within 10 RFLP-defined haplogroups: H, I, J, K, M, T, U, V, W and X (Torroni *et al.*, 1994; 1996).

RFLP and sequencing analyses were then carried out for the same data set so that RFLP-haplogroups could be correlated with control region sequences (Francalacci *et al.*, 1996; Torroni *et al.*, 1996). It was expected that, if the haplogroup classification was correct the control region sequences within the same haplogroup should harbour unique and monophyletic set of nucleotide polymorphisms. Largely, the segregating sites shared among control region sequences classified into the same RFLP-defined haplogroup matched those sites that identified the clusters observed by Richards et al. (1996) –e.g. hg K (as defined by RFLP) control sequences were characterised by sites 224 and 311, which in the RMN of Fig. I.11 define group 4. That is, the correlation between RFLP- and control region- classification was clear-cut. Since then, the alphabetical nomenclature proposed by Torroni et al. (1996) has been adopted.

Using all RFLP and control region sequence data jointly resulted in a more robust phylogeny, with most of the reticulations resolved and some of clusters broken into

further subdivision –yet most of the earlier clusters and general topology survived (Richards *et al.*, 1998; Macaulay *et al.*, 1999a; Sykes, 1999). Although the complete European phylogeny clearly showed 12 haplogroups or lineages (Macaulay *et al.*, 1999a) (see Fig. I.12), given that seven of them were believed to account for ~95% of the haplotypes observed in Europe (Richards *et al.*, 1998), these ancestral lineages have been referred to in a popular text as the 'seven daughters of Eve' (Sykes, 2001).

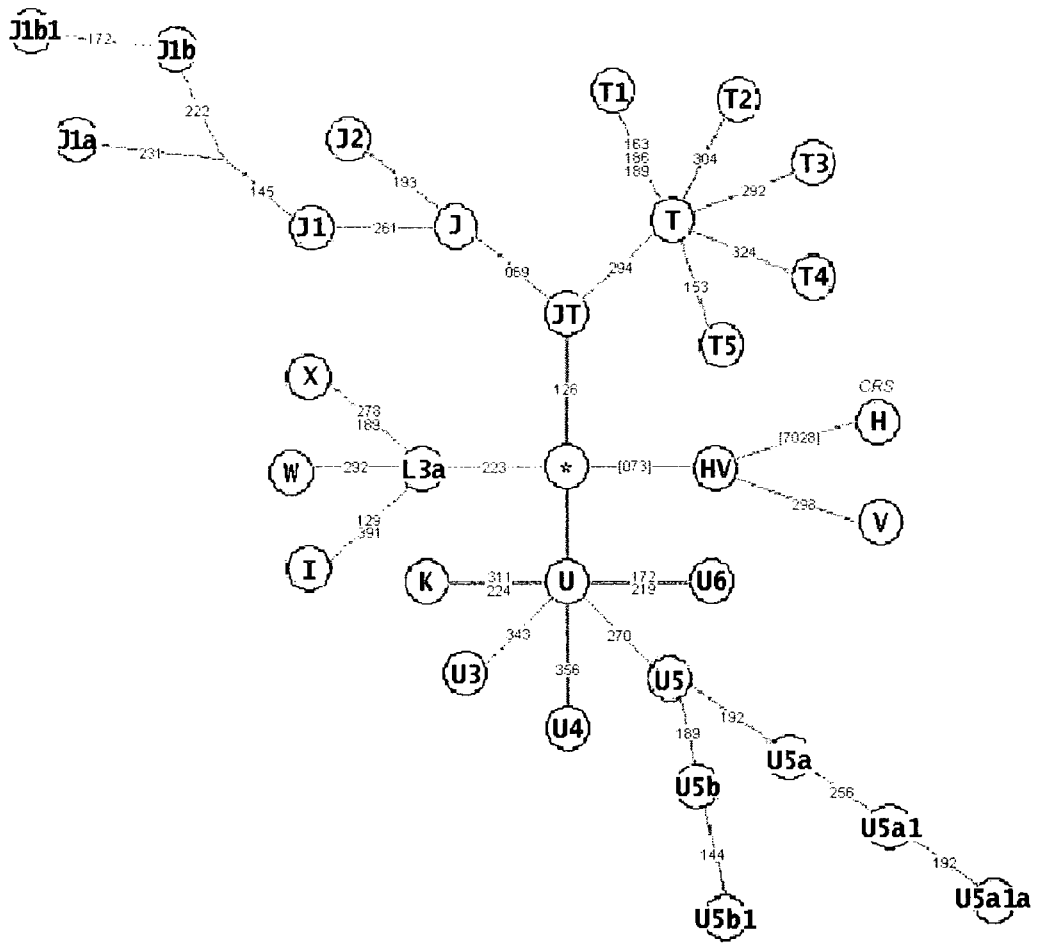


Figure I.12 Schematic tree of European mitochondrial variation. The node marked CRS corresponds to the Cambridge Reference Sequence (Anderson *et al.*, 1981). The branches are labelled with HVS-I mutations (minus 16000 for clarity). HVS-II and coding region mutations are shown in square brackets. Note that some mutations revert within the cluster (e.g. 192 within U5). U6 is very rare in Europe. L3a is a large clade of African origin, based on the root sequence 223 (Watson *et al.*, 1997) (modified and corrected from Richards *et al.*, 1998).

I.3.2 Dating of mtDNA clusters

Richards et al. (1998) repeated the study including more mtDNA sequences from Near Eastern and southern European populations (n = 942). The age of the clusters was this time calculated using the index ρ (instead of π) which takes into account prior information as it requires the designation of ancestral haplotypes (Richards *et al.*, 1998). Rho estimates the internal diversity of a cluster by pairwise comparisons of the sequences within it and their ancestral (also referred to as founder) haplotype (Forster *et al.*, 1996). As in the original paper (Richards *et al.*, 1996), the internal diversity was converted directly to a divergence date by using a mutation rate of 1 transition per 20,200 years (equivalent to 0.18 mutation/bp/Myr) (Forster *et al.*, 1996) –which is close yet not identical to the one used previously. The estimated ages for each major haplogroup are shown in Table I.3. Age of sub-hgs J1a and T1 were also estimated. These sub-haplogroups were thought to represent Near Eastern lineages, as their founder haplotypes were only present in the Near East (Calafell *et al.*, 1996; Richards *et al.*, 1996). Haplogroup dates that had been previously estimated based on RFLP data are also shown for comparison (Torroni *et al.*, 1994; 1996). As can be seen age of clusters estimated by the different methods do not always coincide, partly due to the differences in mutation rates used.

hg	Freq.	Control region[1]	RFLP[2]
H	>50%	18,000–23,000	22,500–29,500
V	4%	9,500–15,500	16,500–22,000
K	7%	12,500–18,500	13,500–18,000*
U5	7%	46,000–59,000	51,000–67,000
T	~8%	40,500–52,500	8,000–11,000
T1	n/a	9,000#	n/a
I	~2%	26,500–43,500	25,500–33,500*
W	~1%	12,000–25,000	n/a
X	~2%	17,500–30,500	24,500–29,500
J	>11%	24,000–32,000	17,000–23,000
J1a	n/a	3,500–9,500	n/a

Table I.3 Age estimates for the major European haplogroups based on control region and RFLP data. (1) from Richards et al. (1998); based on a control region mutation rate of 18% per million years (myr); and (2) from Torroni et al, (1996) (except those marked *, from Torroni *et al.*, 1994); based on a RFLP mutation rate of 2.2-2.9% per myr. Frequencies of haplogroups are those of Richards *et al.*, 1998. (#) reported in the main text of the paper, no CI given.

Given the divergence date of the major haplogroups, the authors claimed that, overall, the results supported their earlier conclusion of a mostly Upper Paleolithic (rather than Neolithic) origin of the European mtDNA gene pool (Richards *et al.*, 1996; 1998). They argued that all the clusters, although to a different extent, underwent expansions during '*European prehistory*' (sic) and proposed that post-glacial dispersals might account for this. A southwest-northeast expansion from a southern refuge was also proposed by Torroni et al. (1998; 2001). This was based on the high frequency and high diversity of hg V in the Iberia Peninsula, which, as mentioned earlier, is believed to have been the location of a southern refuge during the LGM. Dating of the hg V diversity placed its origin at around 13,000 YBP (Torroni *et al.*, 1998; 2001), by this time communities sheltering in refugia are supposed to have repopulated north Europe. Therefore hg V would be the mtDNA marker for the south-north expansion. However, several specific aspects of this result have been questioned (Izagirre & de la Rúa, 1999; Simoni *et al.*, 2000b) (see sections I.3.3 and I.4).

The main criticism of Richards and colleagues' analyses was the idea that the age of haplogroups and expansions can be associated. The dating of European lineages to Paleolithic times should not imply that these lineages *arrived* in Europe at those times. As Barbujani et al. (1998) illustrated '*if Europeans established a population in Mars their common mitochondrial ancestor would be Paleolithic, but it would not be wise to infer from that a Paleolithic colonization of Mars*'. In general, the average coalescent time for two sequences sampled from two diverging populations should be older than the time of divergence. Unless a group colonizing a new territory passes through a strong and long-lasting bottleneck part of its initial diversity will be maintained (Nei, 1975; Tajima 1983). If there was already diversity within the cluster *before* the founder event, then estimates of the arrival dates would be too old, unless this is taking into account.

Sykes (1999) acknowledged this issue and proposed that the way to resolve it was to identify the haplotypes (within each cluster) that were already present in the source population, so that the pre-existing diversity (for each cluster) could be subtracted. It was claimed that the ages of the clusters re-calculated in this manner were substantially different (see Fig. I.13) –however, no data or source of reference is given for this

statement. (It is presumed this data was obtained as part of a posterior paper by the same group; Richards *et al.*, 2000). In any case, it is noticeable that, with the exception of hg U5 and V –which remained at their initial estimated age– all the clusters moved forward, most of them to the period ~10-15,000 YBP. Haplogroup X, the curious haplogroup shared between Europeans and Amerindians, remains at ~20,000 YBP. In particular, now, hg J clearly showed Neolithic origin.

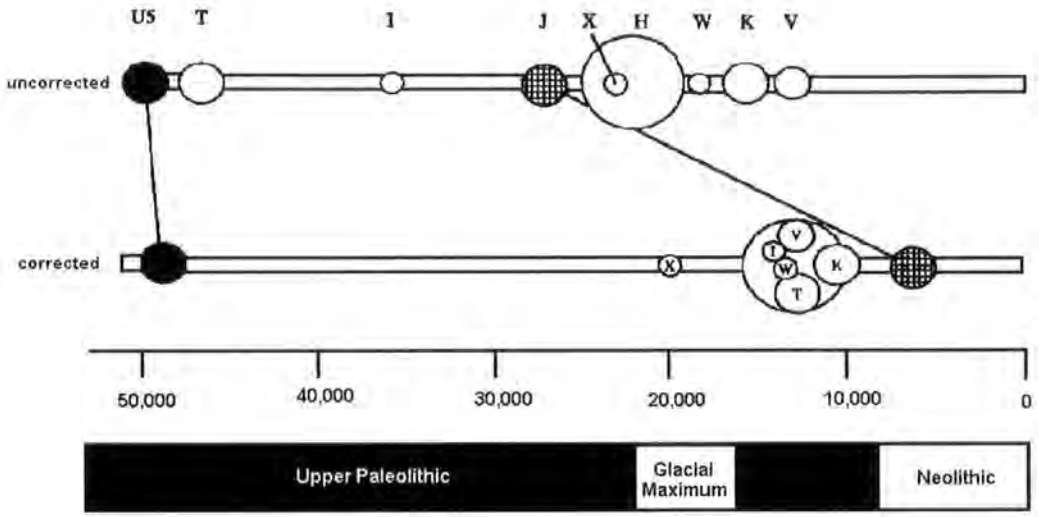


Figure I.13 Uncorrected (Richards *et al.*, 1998) and corrected divergence times for the major European clusters. Areas of circles are proportional to cluster frequencies. NOTE– Glacial Maximum refers to the Last Glacial Maximum (taken from Sykes, 1999).

A much more complex scenario emerges from the comprehensive analyses by Richards et al. (2000). Their aim was to identify the principal lineages that have entered Europe (from the Near East) and to date the times of their entry. This would allow quantification of the contribution that the main episodes of new settlement during European prehistory had made to the modern DNA pool. Most of the major western Eurasian haplotypes were observed in the Near Eastern populations at frequencies $\geqslant$1%. The principal exception to this was hg V, which seemed to have expanded within Europe ~13,000 YBP (Torroni *et al.*, 1998; 2001). In addition, although hg U5 occurred at ~2% in the Near East, its phylogeography seemed to suggest that it evolved within Europe during the past 50,000 YBP (Richards *et al.*, 2000). This would be in agreement with the fact that the expansion dates for hg U5 and hg V were unmodified when accounting for diversity in the alledged source population (see above).

As interpreted in previous studies (Richards *et al.*, 1996; 1998), two major founders are associated with the Neolithic (all hg J and the sub-hg T1) but still the greatest impact on the modern mtDNA pool seems to have been migration during the late Upper Paleolithic, as several clusters were associated it with it (hg T, sub-hg T2, hg K and a sub-group of hg H). In addition, some other haplogroups have earlier dates (hg H, hg U4 and hg I) (see Fig. I.14).

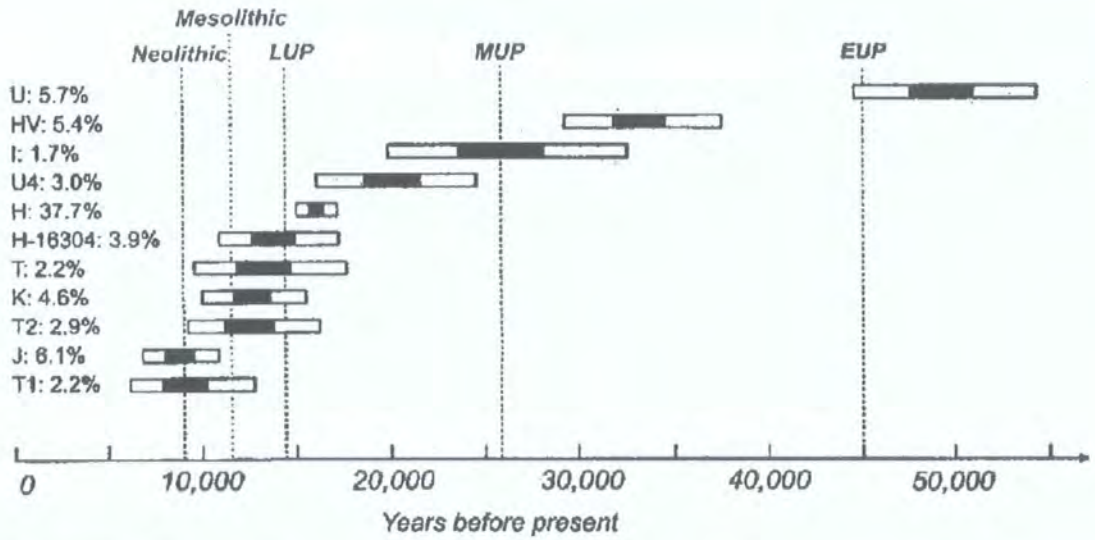


Figure I.14 Age ranges of the major Near Eastern founder clusters. The proportion of sequences in each cluster is indicated. The 95% (or 50%) 'credible region' for the age estimates of each cluster are shown by white (or black) bars (taken from Richards *et al.*, 2000)

It should be noted however, that the study assumed that the Near East was *the* source region for most of the genetic variation extant in Europe. Whereas this is widely accepted for the Neolithic, it is much less secure for earlier times. A Near Eastern origin for the Aurignacian industry, believed to represent the spread of the Early Upper Paleolithic (see section I.1.2), has been proposed (Mellars, 1992), although it is not widely accepted (Clark & Lindly, 1991). Other significant dispersals may have originated in northern Caucasus or eastern Europe (Soffer, 1987; see section I.1.3).

In any case, regarding the more recent expansion, it would seem that hg J has had its origin in the Near East and expanded to Europe with the Neolithic farmers. However, this does not mean that *all* Neolithic farmers were composed exclusively of hg J sequences. Sub-hg T1 has also Neolithic origin, and sub-hgs of K and H have been estimated to be recent as well. Thus, the total Neolithic contribution to the European

mtDNA pool would add up to ⩾20%. This value is not far from the 28% accounted for in the first PC obtained by Cavalli-Sforza et al. (1994). Under this view, the synthetic maps seem compatible with the mtDNA scenario in Europe.

I.3.3 Phylogeography of mtDNA

The geographic distribution of each of the major mtDNA clusters has also been analyzed (Richards *et al.*, 1996; 1998). With a few exceptions, most haplogroups proved to be widely distributed in Europe (see Chapter IV, section I). As opposed to nuclear-encoded protein polymorphisms (Sokal & Menozzi, 1982; Sokal *et al.*, 1991; Cavalli-Sforza *et al.*, 1993) mtDNA did not seem to show clinal distribution along Europe. To better understand this, patterns of mtDNA variation in Europe were analysed by spatial correlation analyses (Simoni *et al.*, 2000b). These analyses compare data (in this case DNA) within arbitrary space intervals. A variable is correlated positively or negatively if its value at a given point in space is associated with its values at other locations. One of the indexes (autocorrelation index for DNA analyses (AIDA), Bertorelle & Barbujani, 1995), based on comparisons of sequences, showed that the overall pattern of mtDNA diversity was poorly structured in Europe (Simoni *et al.*, 2000a; 2000b). However, separate analyses of northern and southern Europe pointed to two different patterns: northern Europe showed only moderate differences, whereas a cline was apparent around the Mediterranean Sea. They argued that this might be due to the faster spread of the Neolithic along southern Europe (see section I.1.4). The other index, (a classical autocorrelation approach: spatial autocorrelation analysis program (SAAP), Sokal & Oden, 1978) based on frequencies of haplogroups rather than sequences, showed very poor pattering. However, Torroni et al. (2000) argued that this result was flawed due to erroneous haplogroup assignment. For the haplogroup assignment the authors used the algorithm proposed by Richards et al. (1998) which, ironically, as noted in the response by Simoni et al. (2000c), was described inconsistently in the original paper. Whereas in Table I of Richards et al. (1998) was stated that hg X is defined by segregating sites 223 and 278, Figure I also includes segregating site 189.

In addition, Simoni et al. (2000b) rejected the possibility of a post-glacial expansion from Iberia into northern Europe, with hg V being its mtDNA marker. The spatial

correlation indexes showed that hg V was not clinally distributed along a transect corresponding to the presumed expansion route. Although based on hg frequencies, this result should still be considered valid, since hg V was not affected by the haplogroup misclassification (Simoni *et al.*, 2000c).

Given that mtDNA did not clearly identify clinal patterns in Europe, the suspicion was raised that protein polymorphisms (such as those analyzed by Cavalli-Sforza et al., 1994) may fail to represent the underlying DNA diversity. However, spatial autocorrelation indices also showed an east-west clinal distribution of nuclear gene frequencies (microsatellite loci) (Chikhi *et al.*, 1998a; 1998b). Dates of population's splits allowed the location in time of the evolutionary phenomena that led to the establishment of these gradients. As opposed to mtDNA, all estimates of times were below 10,000 years, suggesting that the European gene pool was mostly derived from a Neolithic ancestry.

Richards et al. (2002) argued that the absence of mtDNA clines might have been due to a lack of sufficient data for the analyses, rather than to some quirk of the marker itself. To test this they performed PC analyses based on haplogroup frequencies from 10 geographical regions, comprising >4,500 mtDNA sequences from Europe, Near East and North Caucasus. The first principal component (PC) (explaining 51% of the variation), was primarily east to west, separating European from the Near Eastern populations: Iraq was at one pole of the axis and the Basque Country at the other. The central and eastern Mediterranean populations and the southeastern populations of Europe plotted with the other Europeans and also showed affinities with the Near Easterners. The main haplogroups contributing to this pattern were hg H and hg U5 concentrated at the European pole and hg U1 at the Near Eastern pole (see Fig. I.15).

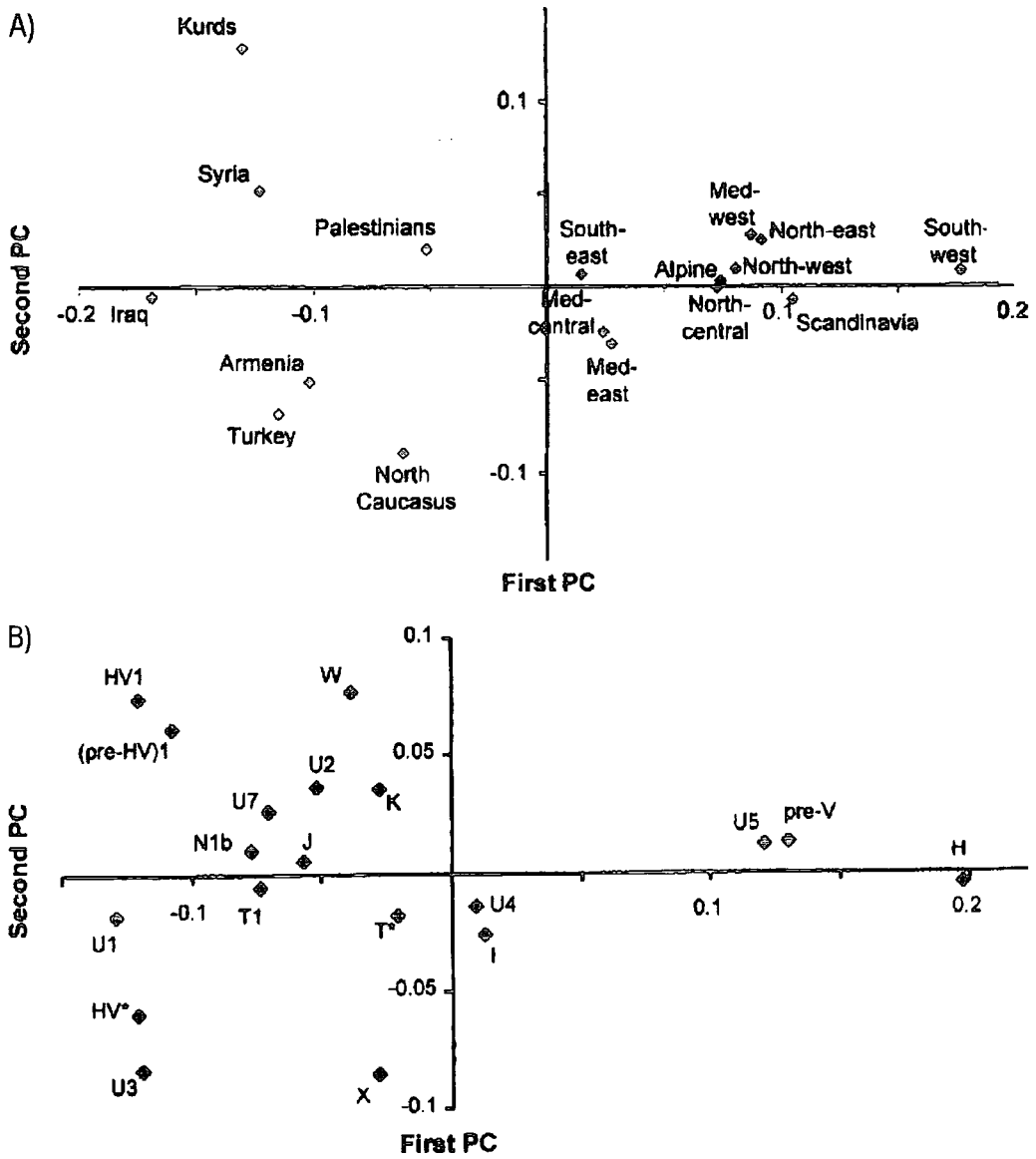


Figure 1.15 PC analyses of mtDNA in Europe and the Near East. (**A**) Region-based PC analyses of mtDNA haplogroup profiles. (Med) Mediterranean. (**B**) Plot of the contribution of each haplogroup to the first and second PC in the analysis shown in (A). NOTE– some of the haplogroups (e.g. pre-V) are RFLP-defined (taken from Richards *et al.*, 2002)

The analyses seemed to indicate that mtDNA also had an (approximately) southeast-northwest geographic pattern, similar to those observed for classical markers. However, the authors believed that this pattern is unlikely to represent the contribution of Near Eastern farmers to the mtDNA pool. Haplogroup frequencies indicated similarity between Europeans and Near Easterners primarily in southeastern Europe and along the Mediterranean, whereas archaeological evidence, they argued, would point to the main expansion of agriculture being into central Europe. However, early archaeological data

indicated that agriculture expanded first, quickly to the north of the Mediterranean through Greece and the Balkans, and then more slowly into the rest of Europe (Clark, 1965a; 1965b) (see section I.1.4). This view would be in agreement with Richards and colleagues' findings (2002). In addition, a previous study also detected a mtDNA cline around the Mediterranean which was interpreted in the same way (Simoni *et al.*, 2000). However, what is still difficult to reconcile with the observed mtDNA pattern being the result of Neolithic spread is that the haplogroups accounting for the first PC are not those believed to have Neolithic origin (Richards *et al.*, 2000) (see section I.3.2). On the contrary, the mtDNA pattern (as depicted by the first PC) seemed to be due to haplogroups that previous analyses assessed as having originated *within* Europe or spread into Europe in the Upper Paleolithic such as hg U5 and hg V. The authors conclude that the mtDNA data might be reflecting the superimposition of several processes occurring at different periods in history and prehistory.

It is important to bear in mind, as stated by several authors (e.g. Richards *et al.*, 2000; Adcock *et al.*, 2001), that all these results are inferences based on the *present-day* mtDNA pool. Contemporary DNA sequences provide only indirect evidence of the historical processes that have formed them over long periods of time. Only the direct analysis of ancient DNA will fully confirm a common past, as other processes might cloud the true story.

I.4 Ancient DNA analyses

The prospects of ancient DNA (aDNA) work in understanding evolutionary and historical processes are obvious. Unfortunately, its study poses numerous technical pitfalls, as only few broken and modified molecules are available to study (Pääbo, 1989; Lindahl, 1993; Gilbert *et al.*, 2003a) (see Chapters III and IV). The development of the PCR technique (Mullis & Faloona, 1987) however, greatly contributed to the expansion of ancient DNA as a field.

Ancient DNA analysis has given insight into the phylogenies of extinct species (e.g. geese, Fleischer *et al.*, 2000; wolves, Thomas *et al.*, 1989; bovids, Lalueza-Fox *et al.*, 2002;wrens, Cooper 1994; dogs, Leonard *et al.*, 2002 and bears, Orlando *et al.*, 2002), the evolution and origin of diseases and pathogens (e.g lepra, Haas *et al.*, 2000;

tuberculosis, Baron *et al.*, 1996; Zink *et al.*, 2001; plague, Drancourt *et al.*, 1998; Zink *et al.*, 2002 and chagas, Ferreira *et al.*, 2000) and the domestication of cereals and animals in ancient times (e.g. sorghum, Deakin *et al.*, 1998; maize, Goloubinoff *et al.*, 1993; wheat, Brown *et al.*, 1994; cattle, Bailey *et al.*, 1996; goat, Kahila Bar-Gal *et al.*, 2002). However, with regards to human aDNA, the situation is rather different, since the recovery of ancient DNA has so far provided only a few significant outcomes.

It was Pääbo (1985) who first reported having cloned a fragment of DNA from a 2,400 year-old mummy, and some years later, courtesy of the PCR reaction, directly amplified mtDNA from an 7,000 year-old brain (Pääbo *et al.*, 1988). The body had been recovered from a peat bog in Florida and it is believed that the anoxic conditions meant that the DNA was remarkably well-preserved. Since then, a large number of scientific papers have been published in the human aDNA field. However, many –if not most– of them were merely reports of the successful retrieval of aDNA from one or few ancient specimens (e.g. the Tyrolean Ice man, Handt *et al.*, 1994b; the Cheddar man, unpublished, but see Sykes, 2001) which did not provide much understanding of human history. The great exception of this is, obviously, the analysis of the Neanderthal specimens (Krings *et al.*, 1997; 1999; 2000; Ovchinnikov *et al.*, 2000) and their contribution to resolve the debate about the origins of anatomically modern humans (although this is still controversial).

Although some work has been attempted on single locus genes (mainly sex determination genes, e.g. Faerman *et al.*, 1995; Dudar *et al.*, 1997; Cipollaro *et al.*, 1998; Matheson & Loy, 2001; but also STRs, Zierdt *et al.*, 1996; ABO group, Hummel *et al.*, 2002; ß-globin, Béraud-Colomb *et al.*, 1995; Filon *et al.*, 1995 and Y-chromosome, Hummel & Herrmann, 1991) aDNA studies have mostly focused on mtDNA. The simple reason is that, unlike nuclear DNA, mtDNA is multicopy (there are >800 mtDNA molecules in each human cell, Robin & Wong, 1988) and therefore is much easier to extract and amplify from ancient samples. Nevertheless, the DNA extraction success rate is rarely >75% –and if it is, should be a warning for possible contamination (Kolman & Tuross, 2000). This, together with the great difficulties of authentication and the uniqueness of archaeological specimens (which implies that only few samples are available for study) meant that researchers did not endeavour in human

population-level studies. In fact, in 14 years of research only a dozen studies have attempted to analyse more than 20 archaeological samples (see below). However, the fact that these do exist shows that, although extremely time-consuming, expensive and discouraging there is scope for human aDNA work. It should also be noted that, paradoxically, a population-level study may lend itself to authenticate the results (Hauswirth *et al.*, 1994; e.g. see this study). Diversity of the data is an internal element of self-validation, as (reproducible) contamination with several sources is unlikely.

The first colonization of the Americas is an interesting case for the application of aDNA analyses. Ancient DNA has contributed to our understanding of the time and number of migrations to the New World (Stone & Stoneking, 1993; Parr *et al.*, 1996) and the continuity between ancient and modern native populations (González-Oliver *et al.*, 2001). Although at different frequencies, the majority of ancient mtDNA haplotypes fall into one of the four original Amerindian haplotypes, namely A, B, C and D (see section I.2.4). These hg are also found in East Asia and Siberia, in agreement with them representing migratory event(s) through the Bering straight around 17-34,000 YBP (Torroni *et al.*, 1993). However, for the more recently found fifth haplotype in modern Amerindians, namely hg X, the migration route was not as clear, since this hg had not been found in any Asian population but is frequent in European populations (Brown *et al.*, 1998). Diverse theories were proposed to explain the presence of hg X in Native Americans. Stanford (see Fortune, 2002) –based on lithic similarities– suggested the possibility that early Europeans colonized the Americas via a trans-Atlantic route. Others proposed that hg X had been brought by recent European colonizations (Brown *et al.*, 1998). The issue has been resolved by modern and ancient mtDNA research. On one hand, recent mtDNA analyses reported the presence of hg X in Altaian populations of south Siberia (Derenko *et al.*, 2001) and on the other hand, ancient DNA analyses have shown that hg X was already present in native American communities before admixture with modern Europeans (Stone & Stoneking, 1998; Malhi & Smith, 2002).

Ueda's team has been carrying out aDNA studies of Asian populations (Kurosaki *et al.*, 1993; Oota *et al.*, 1995; 1999; 2001; Wang *et al.*, 2000). The most interesting study analyses a 2,500-year-old and a 2,000-year-old population to investigate the genetic continuity between ancient and modern Chinese (Wang *et al.*, 2000). A 185-bp fragment

of HVS-I was amplified in 34 samples from the 2,500-year-old site and analysed together with previous data of 13 samples from the 2,000-year-old site for which the same length sequences were available (Oota *et al.*, 1999). In a RMN, constructed including these two ancient populations and a sample of modern Chinese, 6 clusters were identified, named I-VI (classification according to the established nomenclature was not used due to lack of resolution). Interestingly, the most frequent haplotype (65%) among the 2,500 year-old population was observed at a lower frequency (13%) in the 2,000 year-old population but not observed among the modern Chinese population. Hence, a notably decrease in haplotype frequency, leading to extinction, seems to have occurred in the Chinese population. In addition, a distortion of cluster frequencies over the time and a decrease in nucleotide diversity was also observed.

In Europe, very few population-level studies have been carried out. Gerstenberger (2002) investigated wedding patterns in a Medieval German community by mtDNA and Y-chromosome analyses. Higher mtDNA than Y-chromosome diversity suggested that patrilocality was practiced. An earlier study attempted amplification of one STR locus in a Medieval population (no date was given) (Zierdt *et al.*, 1996).

A remarkably large and probably the most relevant study of European aDNA has been carried out by Izagirre and de la Rúa (1999) in an RFLP analysis of four Basque prehistoric sites. The authors tested the hypothesis proposed by Torroni et al. (1998) in which hg V is considered the mtDNA marker for a major Mesolithic expansion from south-western Europe. This hypothesis had been based on the estimated age of hg V (~13,000 YBP) and its considerably high frequency in populations of the Iberia Peninsula (including the Basques) from where it declined towards both the north-east (and south), as would be predicted if the hg V frequency was the result of a post-glacial expansion from its homeland –a south-western refuge. If so, the frequency of hg V in ancient Basques should be also high. Interestingly, among the >120 Neolithic Basques (3,500 to 5,000 YBP) samples analysed not a single hg V was observed, which conflicts and weakens Torroni et al. (1998) hypothesis. To explain the differences between the modern and ancient Basque population, the authors speculated that either the age of hg V might be more recent than estimated by Torroni et al. (1998) (due to inaccurate estimation of the mutation rate, see section I.2.3) or that hg V arrived to the Basque

region <3,500 year ago. In any case, the results obtained by Izagirre and de la Rúa (1999) strongly points to the possible limitations of the extant DNA data to infer past events and highlights the importance of aDNA studies in the analyses of human history and evolution.

I.5 Britain

Despite the lasting cultural and linguistic impact, there is considerable disagreement regarding the scale of the Anglo-Saxon settlement in Britain after the withdrawal of the Roman army. Estimates vary from a take-over by a small elite to a migration of up to perhaps hundred thousand people (Härke, 2002) (see Chapter V). There is evidence of genetic clines in Britain (Falsetti & Sokal, 1993) as one might expect had a major immigration centred on the south-east taken place. However, whether this represents Saxon immigration or earlier events is unclear.

Ancient DNA analysis of Anglo-Saxon settlements is believed to have the potential to help clarify this issue and has therefore been the focus of previous research. However, early work including samples from different locations and periods yielded very few results which prompted the authors to suggest that the British climate and soil were not appropriate for bone preservation (Richards *et al.*, 1993; 1995). Subsequent attempts on early Anglo-Saxon settlements yielded only half a dozen mtDNA sequences (Richards *et al.*, 1999 –work carried out in 1995). Hence, no population-level ancient DNA study has hitherto been successfully carried out in the British Isles.

I.6 Objectives

The main objective of this research is to contribute to our understanding of the history of the British population from a genetic perspective, by the direct study of ancient communities using matrilineal markers.

The first challenge of this work will be to obtain DNA from British archaeological material in relatively large number for a population-level study. As mentioned, DNA from ancient material is scarce and highly damaged so that its extraction requires extremely sensitive protocols. It will be, therefore, necessary to extensively optimise the

methods and to gain the expertise required to extract DNA from poorly preserved hard tissues (see Chapter III). Part of this serves not only to retrieve aDNA from the archaeological material but, above all to authenticate it.

Given a sample size of authenticated sequences sufficient to assess population-level diversity, the following hypotheses and questions will be addressed:

1) Matrilineal continuity between Anglo-Saxon communities and English modern population would be expected, provided females migrated with the Anglo-Saxon invaders (see Chapter IV);
2) Genetic diversity of the population of Britain would have increased since Anglo-Saxon times due to admixture with subsequent invading populations, such as Vikings and Normans (see Chapter IV);
3) Mitochondrial DNA haplotypes observed in the ancient population of Britain will be expected to cluster among the major European haplogroups? (see Chapter IV);
4) Genetic composition of settlements from different historical periods, such as pre-Saxon (4^{th} century), early Saxon (5^{th}-7^{th} century) and late Saxon (9^{th}-11^{th} century) will differ due to the influence of the respective Anglo-Saxon and Viking invasions (see Chapter V);
5) Given the proposed pattern of post-glacial expansion of human populations, the genetic composition of early populations of Britain will reflect a re-colonization process originated in southern refugia, in particular south-western France (see Chapter V).

Chapter II - Materials and methods

II.1 MATERIALS

II.1.1 Archaeological sites and samples

Five archaeological sites dated from 4th to 11th century were studied. The sites, located mainly in the southern and eastern coast of England, correspond to the ancient settlements of: Newarke Street at Leicester; Dover, Buckland at Kent; Grove Farm, Market Lavington at Salisbury; Norton at Cleveland and Farmer's Avenue, Castle Mall at Norwich. Figure II.1 shows the geographical location of the sites.

The sites were excavated by Leicester Archaeological Unit (LAU), Canterbury Archaeological Trust Ltd; Kent, Wessex Archaeology; Salisbury, Cleveland County Council Archaeological Section; Cleveland and Norfolk Archaeological Unit (NAU); Norwich, respectively, which gave official permission to access to their archaeological material.

The historical period of each cemetery was determined by the respective archaeological units on the basis of the archaeological finds excavated. Newarke Street at Leicester was assessed as Romano-British period (Cooper, 1996) whereas all the other sites were identified as Anglo-Saxon period (Evison, 1987; Sherlock & Welch, 1992; Parfitt, 1995; Williams & Newman, in prep.; Popescu, in prep.). Details of the sites are given bellow.

When possible, sex and age of the individuals were assessed by osteological and anthropometrical methods. The sex of a number of skeletal remains could not be determined as these techniques do not allow unequivocal sexing of immature and juvenile individuals (Brothwell, 1981). Gender, that is, the 'sex' assigned to an individual based on the associated grave-finds was also determined. Sex and gender may or may not match. Figures II.2 and II.3 show examples of the grave-goods found in a male and female grave.

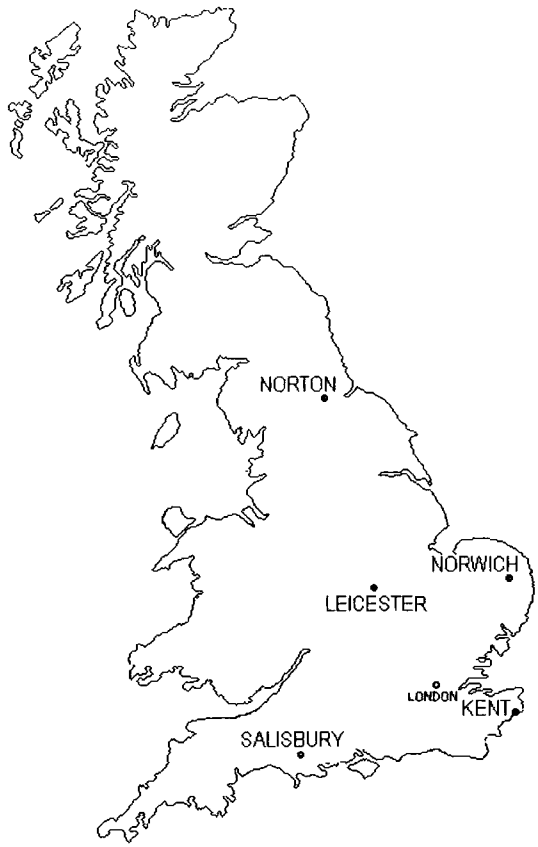


Figure II.1 Geographical location of the archaeological sites used for this study.

Samples.

The archaeological material analysed consisted of dental pieces belonging to the human skeletal remains exhumed from the above mentioned sites. Due to the archaeological value of the material (and the destructive nature of the analyses) only two teeth per individual were given for the study. Samples were mostly molars, premolars or otherwise the largest tooth available. At macroscopic level, teeth were in good condition, though this varied from site to site and from specimen to specimen. Figure II.4 shows examples of the dental samples analysed.

Figure II.2 Example of a male grave (number 346 from Dover, Buckland at Kent), showing the skeletal remains and his grave-goods: a sword, a long spearhead and an axe below his right foot. The individual is believed to have been a high status warrior. The skeleton was sexed as male on anatomical basis (Parfitt, 1995).

Figure II.3 Example of a female grave (number 250 from Dover, Buckland at Kent), showing the skeletal remains and her grave-goods: a gold pendant, beads, an ivory ring for a purse by her left wrist, glass vessels below her feet, an iron weaving batten by her left ankle. It is believed that the individual was a rich female although data about the anatomical sex was not available.

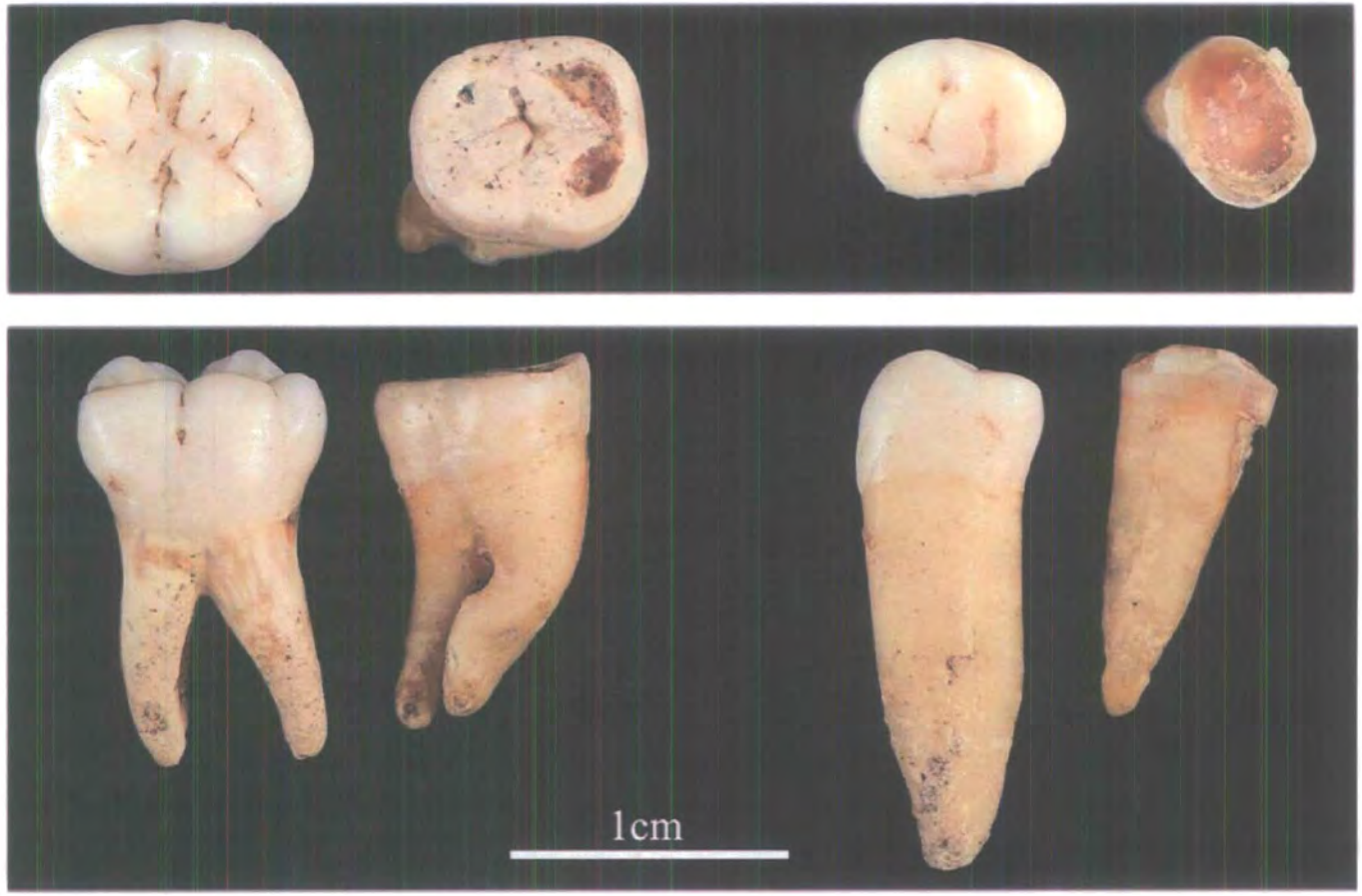


Figure II.4 Dental archaeological material used for DNA analyses. Two molars (left) and premolars (right) and different preservation states are shown. The best preserved teeth available were used for the DNA extraction. Samples in the picture are from the Farmer's Avenue, Castle Mall site at Norwich dated 800-1000 AD.

II.1.1.1 Newarke Street at Leicester

The excavation was carried out by the Leicestershire Archaeological Unit (LAU) in 1993. The site is located at the rear of the Law School in Newarke Street, Leicester, some 125 m. to the south of the Roman and medieval town wall. The area is known to have been within a Roman-British cemetery, evident from many of the burials observed during past building work.

The excavated material consisted of 54 skeletons: 11 males, 12 females, 11 children and 20 unsexed individuals due to incompleteness or fragmentation of the skeletons. Grave pits were rectangular with vertical sides and flat bases with dimensions appropriated for each corpse. Soil conditions were such that timber coffins would not survive physically, nor as a stain, and therefore burials in coffins were deduced from the presence of nails (Cooper, 1996). At least 21 graves contained coffins. Of the 54 skeletal remains excavated, 15 samples were used for the DNA analyses. Table II.1 shows details of sex, age, period and other characteristics of the individuals. Layout of the cemetery and position of their graves is shown in Figure II.5. The lack of deliberately deposited finds (grave-goods) caused some problem with the dating of the cemetery. Yet it was concluded that it was short-lived and restricted to the 4th century.

Note – hereinafter this site will be referred to as 'Leicester'.

Grave	Skeleton	Sex	Age	Grave finds	comments	DNA
15.7	538	female	17-25	coin dated 347 AD		R1
15.15	247	female	25-35			R2
15.2	567	male	45+	coffin?	parrier fracture	R3
15.1	574	male	30-35	coffin?	over 15.35 and 15.36	R4
15.19	424	male	45+			R5
15.20	427	female	17-25	coffin?		R6
15.34	602	female	45+	coffin?		R7
15.36	590	?	adult		overlaid by 15.1	R8
15.6	530	male	45+			R9
15.35	593	?	60+		overlaid by 15.1	R10
15.14	310	male	25-35	coffin?		R11
15.8	534	female	25-35	coin dated 270 AD		R12
15.18	425	male	45+	Brooche (c.225 AD)		R13
15.39	283	?	?			R14
15.33	598	male	25-35	coffin?		R15

Table II.1 Details of the skeletal remains used for the DNA analysis from Newarke Street cemetery at Leicester. Age is indicated in years. Last column (DNA) shows their labels during the DNA analyses.

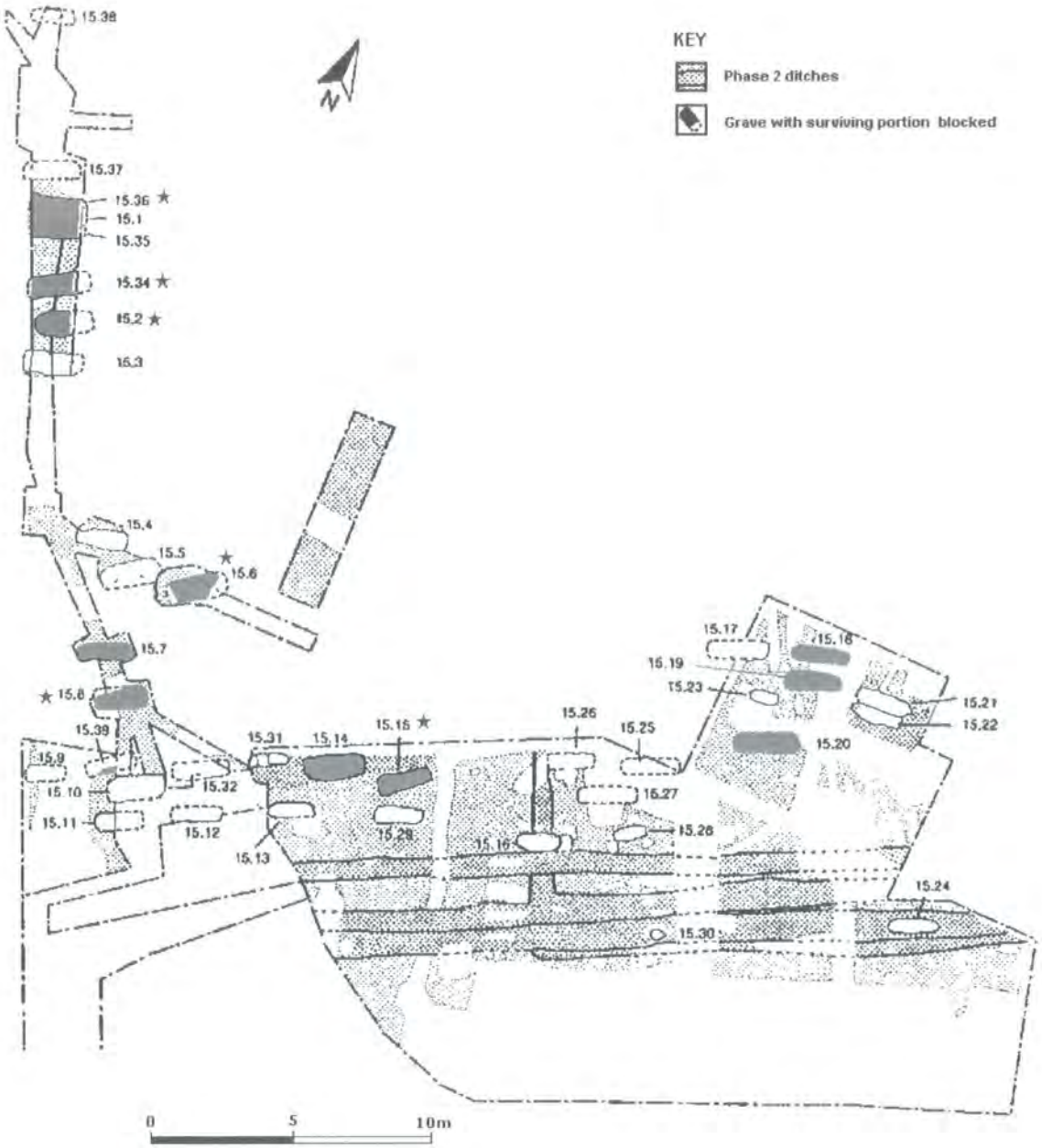


Figure II.5 Layout of the Newarke Street cemetery at Leicester. The skeletal remains used for the DNA analysed are indicated in grey. Skeleton 598 (grave 15.33) was immediately beneath grave 15.3.

II.1.1.2 Norton at Cleveland

The Norton cemetery at Cleveland laid on the Tees valleys. The excavation was carried out between 1983 and 1985 by the Cleveland County Council Archaeology Section. It revealed 117 inhumation burials and 3 cremations. Although the quantity of disturbed skeletal material indicates that several graves might have been eroded, the greater part of the cemetery was successfully recovered by the excavation. Of the 117 total graves excavated, 23 skeletal remains were used for the DNA analyses. Details of sex, age, gender, period and other characteristics of the individuals are given in Table II.2. Layout of the cemetery and position of their graves is shown in Figure II.6.

Dating of the site was based on the archaeological finds excavated and estimated to be from 5th to 6th century. The cemetery at Norton seems to have served a small community and gives a near-complete view of an Early Anglo-Saxon community within the former Bernician kingdom.

Note – hereinafter this site will be referred to as 'Norton'.

Grave	Sex	Gender	Age	Position	Grave-goods	DNA
55	male	male	25-35	extended	spearhead, shield, spear ferrule, knife	N1
94	female	female	25-35	extended	annular brooche, beads, pendant, wrist clasp, knife	N2
63	male	female	25-35	extended	brooches, necklace, ring, bucket, clasp, key, knife	N3
49	female	female	20-30	extended	annular brooche, glas vessel, knife, buckle, pin	N4
36	male	female	15-21	extended	ring, pot	N5
56	female	female	30-40	crouched	annular brooche, beads, buckle, spangles, knife	N6
59	male	female	35-45	extended	ring, wrist clasp, bead, knife	N7
84	female	female	35-45	prone	cruciform & annular brooche, clasp, beads, ring, knife	N8
91	male	n/a	17-25	prone		N9
10	female	female	15-21	extended	beads, glass	N10
66	?	female	10	extended	necklace, matching brooche, ring	N11
71	male	female	15-21	extended	necklace, pendant	N12
65	female	female	30-40	extended	penannular brooche, knife	N13
85	female	female	16-20	extended	annular brooche, buckle, beads, ring	N14
24	male	male	15-21	extended	spearhead, knife,	N15
69	male	male	17-25	extended	spearhead, buckle, beads	N16
18	male	n/a	25-35	extended		N17
68	female	female	35-45	extended	wrist clasps, beads	N18
90	male	female	25-35	extended	brooch, ring, beads, knife	N19
12	male	male	21-25	extended	sperahead, buckle	N20
41	female	female	20-30	crouched	penannular brooches, penannular ring, knif, beads	N41
57	male	female	17-25	crouched	brooch, toilet implement, wrist clasp	N57
120	male	male	25-35	extended	spearhead, bucket	N120

Table II.2 Details of the skeletal remains used for DNA analyses from the Norton cemetery at Cleveland. Grave and skeletal numbers are the same. Age is indicated in years. (DNA) corresponds to the label during the DNA analyses.

Figure II.6 Layout of the Norton cemetery, showing all the burials excavated (except for cremation 119 to the far left). Location of the 23 skeletal remains used for DNA analyses is shown in grey. Samples for which DNA was obtained are further indicated with a star. Doted lines link individuals possibly maternally related (see Chapter IV).

II.1.1.3 Dover, Buckland at Kent

The cemetery laid on the outskirts of Dover, Buckland at Kent. The excavation was carried out in two stages. The first part of the excavation was done by Vera Evison in 1951 (Evison, 1987) and second part in 1894 by the Canterbury Archaeological Trust (yet unpublished). Some 165 and 250 graves were dug up in the first and second excavation, respectively, and was estimated that the complete cemetery could have contained well in excess 500 graves. Dating of the site was based on the archaeological finds excavated. The Buckland cemetery is not one of the very earliest Saxon cemeteries (unless the early graves were destroyed by previous building work in the site) since it flourished mainly in the 6th and 7th centuries. The cemetery was divided into seven periods ranging from 475 to 750 AD, yet most of the graves fall into the 6th century. Of the 250 graves unearthed in the second excavation, 31 skeletal remains were used for the DNA analyses. Details of sex, age, gender, period and other characteristics of the individuals are given in Table II.3. Layout of the cemetery and location of their graves within the cemetery is shown in Figure II.7.

A complete community with men, women and children was buried in the cemetery. It apparently represents the traditional place of a well-established wealthy local community. The wide range and variety of objects found confirm the cosmopolitan nature of the Kentish kingdom during the sixth century.

Note – hereinafter this site will be referred to as 'Buckland'.

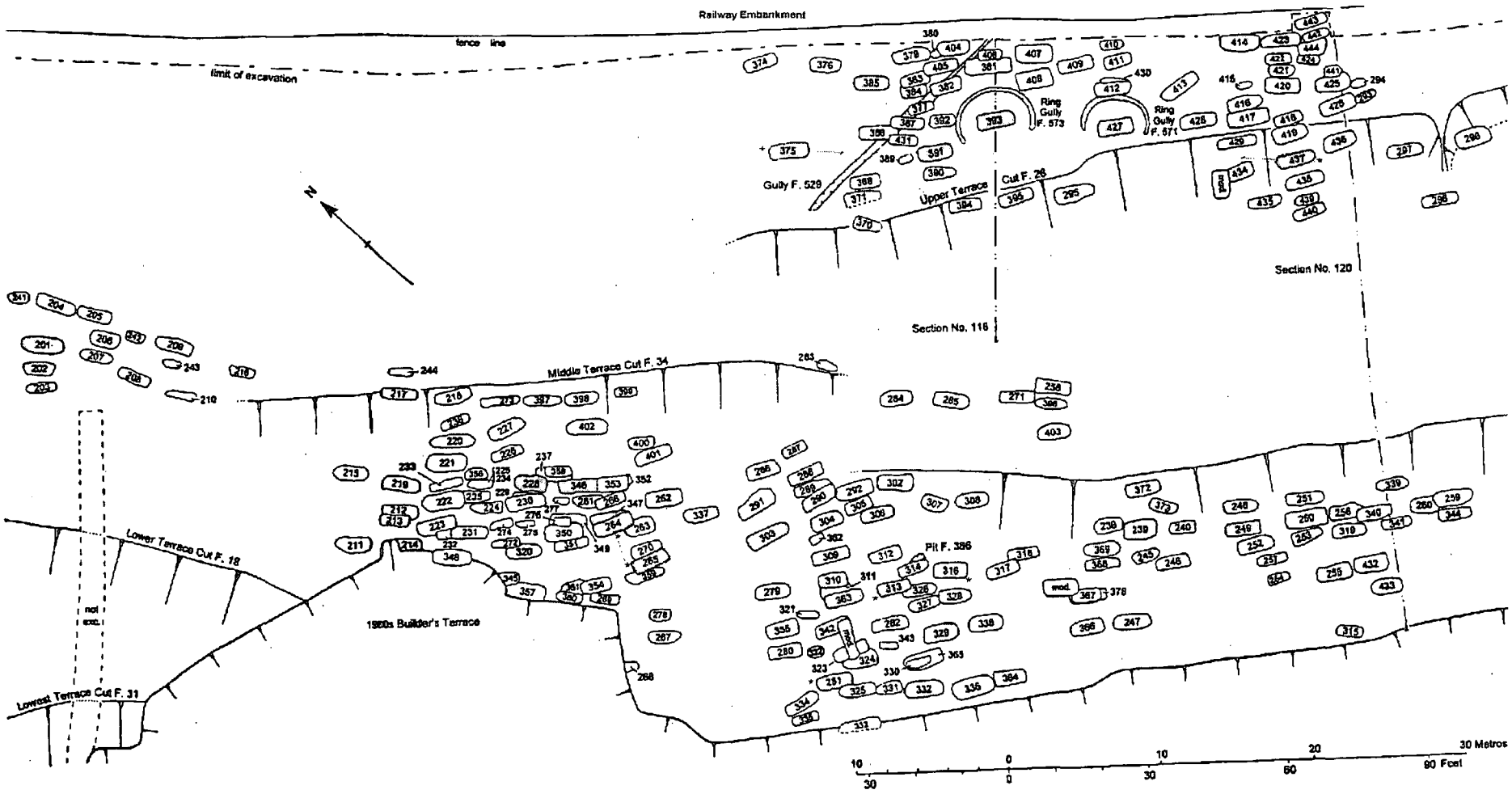


Figure II.7 Layout of the Dover, Buckland cemetery at Kent, showing all the burials unearthed during the second phase of the excavation. Location of the 31 skeletal remains used for DNA analyses is shown in grey. Samples for which DNA was obtained are further indicated with a star. Doted lines link individuals possibly maternally related (see Chapter IV and V).

Grave	Sex	Gender	Age	Phase	DNA
201	male	-	25-30	-	B1
202	female	-	20-25	-	B2
203	female	-	25-30	-	B3
228a	female?	?	15-17	550-640	B4
228b	male?	-	17-20	550-640	B5
239	female	female	30-40	475-525	B6
249	male	male	40-50	525-590	B7
255	female	female	30-35	510-570	B8
264	male	male	40-45	550-640	B9
265b	male?	male	25-30	550-640	B10
281	male?	female	25-30	510-540	B11
303a	?	?	c.3	550-640	B12
303b	male	-	30-35	-	B13
313	female	-	c.40	525-640+	B14
314a	female	female	c.40	-	B15
314b	?	-	7-9	-	B16
316	female	-	65+	-	B17
329	male	-	25-30	-	B18
346	male	male	40+	525-640	B19
347	female?	female	14-16	525-570	B20
369	female	-	30-35	-	B21
375	male	male	25-30	575-640	B22
385	male	-	20-25	-	B23
391a	female	?	35-45	640+	B24
391b	female	female	20-25	525-590	B25
393a	female	?	25-30	-	B26
393b	female	female	35-45	-	B27
414	male	male	35-45	525-570	B28
427a	female	?	30-35	460-570	B29
427b	?	?	20-25	460-570	B30
437	male	male	35-40	510-570	B31

Table II.3 Details of the skeletal remains used for DNA analyses from the Dover, Buckland cemetery at Kent. Grave and skeletal numbers are the same. (Gender) indicates the sex of the individual based on the archaeological finds. Age is indicated in years and phase is AD. (DNA) corresponds to the label during the DNA analyses.

II.1.1.4 Grove Farm, Market Lavington at Salisbury

The village of market Lavington, Wiltshire is located across a small ridge within the bench which runs parallel the northern chalk escarpment of Salisbury. The excavation was carried out by Wessex Archaeology during 1986. A total of 42 inhumation burials were unearthed, which are believed to represent only a sample of the total population of the cemetery. Forty-one of the 42 graves contained skeletal remains. None of the skeletons was complete and the condition of the bone was very poor. There were 27 adults, 12 juvenile and 2 could not be aged. Of the adults, 10 were male and 9 female

but the remaining of the individuals could not be sexed. Figure II.8 shows the layout of the cemetery.

Plenty of archaeological evidence suggests that market Lavington was an Anglo-Saxon settlement. Some of the ditches were datable to the early Saxon period and the age of the cemetery from mid 6th to 7th century (Mepham, pers.comm.).

Of the 41 exhumed skeletons, 28 dental samples were used for the DNA analyses. Location of their graves is further indicated in Figure II.8. Table II.4 shows details of sex, age and other characteristics of the individuals.

Note – hereinafter this site will be referred to as 'Lavington'.

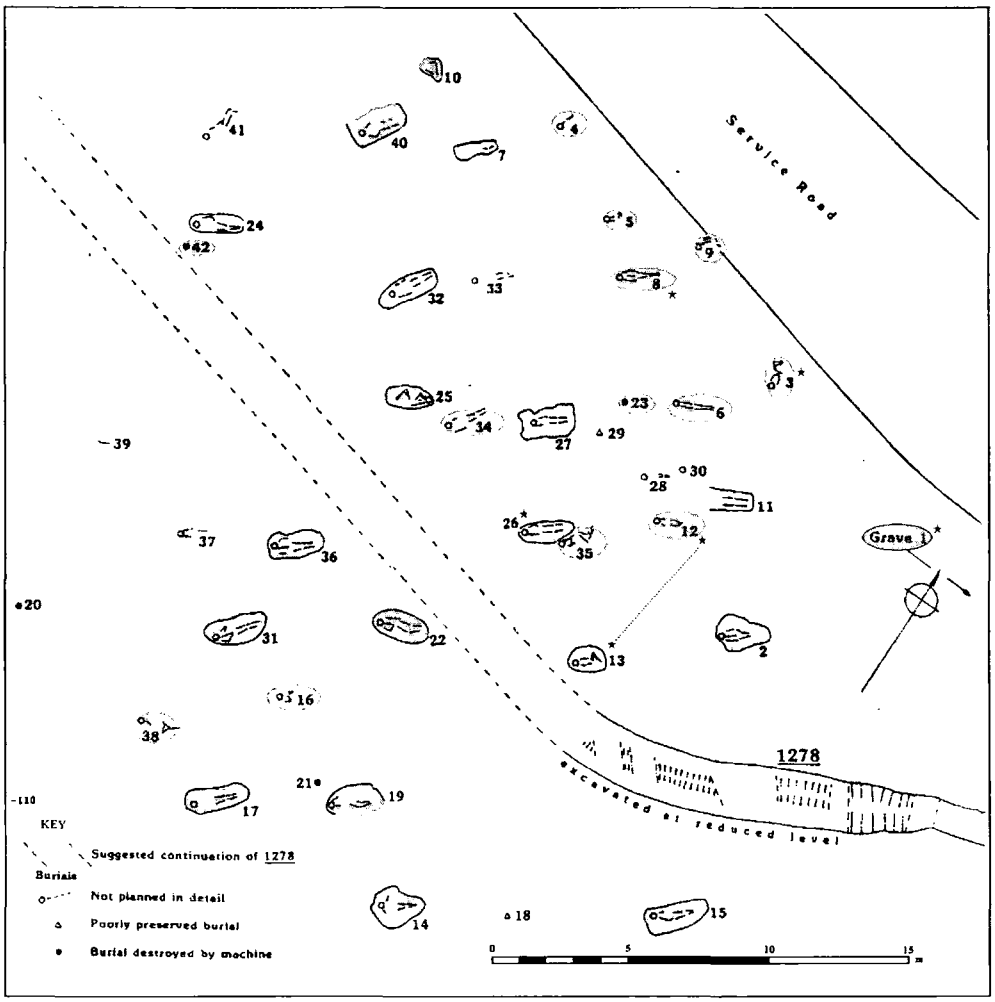


Figure II.8 Layout of the cemetery at Grove Farm, Market Lavington at Salisbury (showing all the graves except for grave 1). The skeletal remains used for the DNA analyses are indicated in grey. Samples for wich DNA was obtained are further indicated with a star.

Grave	Skeleton	Sex	Age	Grave goods	DNA
1	1004	female	young adult		L1
2	1005	?female	adult	knife	L2
3	1006	n/a	teen	spearhead, knife	L3
4	1016	?female	adult	matching brooches, beads, vessel	L4
5	1021	n/a	3-5	beads	L5
6	1022	n/a	c.15	spearhead, shield	L6
8	1029	female	young adult	golded brooch, knife, beads	L7
9	1034	male	young adult	spearhead	L8
10	1039	n/a	adult	awl	L9
12	1046	n/a	6-10	knife	L10
13	1049	n/a	teen		L11
14	1144	n/a	adult	knife	L12
15	1145	?female	adult	knife	L13
16	1155	?male	adult	beads	L14
17	1156	?male	adult	spearhead, shield	L15
19	1164	?female	adult		L16
22	1182	male	old adult	knife	L17
23	1185	?female	old adult	knife, pins	L18
25	1187	male	adult		L19
26	1192	?female	adult	brooches, cosmetic pick?, knife, pins	L20
27	1195	n/a	immature	spearhead, knife	L21
31	1207	male	adult	spearhead	L22
34	1217	?male	adult	spearhead, tweezers	L23
35	1219	male	old adult	spearhead, shields, buckle	L24
36	1222	female	young adult	pins, buckle	L25
38	1228	n/a	10-14	knife, vessel	L26
40	1238	n/a	adult	knife	L27
42	1250	n/a	3-5		L28

Table II.4 Details of the skeletal remains used for DNA analyses from the Grove Farm, Market Lavington cemetery at Salisbury. Age is indicated in years. (DNA) corresponds to the label during the DNA analyses.

II.1.1.5 Farmer's Avenue, Castle Mall at Norwich

The area surrounding Norwich Castle was known to overlie a substantial part of the pre-Conquest settlement of Norwich, one of the largest towns in England by 1066. A large scale excavation was undertaken by the Norfolk Archaeological Unit between 1987 and 1991. Four cemeteries were unearthed to the south of the outer bailey of the Castle and several others in the northern area of the Medieval city (see Fig. II.9). A Late Saxon cemetery was found beneath the later southern bailey rampart. Ceramics from grave fills suggest that this cemetery may have been established in the late 10th century, with burial continuing into the mid 11th century. Radiocarbon dates however indicated that the burial ground may have originated as earlier as the late 9th century (cal AD 890-1020). A duration of a maximum of 155 years was suggested (Bayliss, pers.comm.).

Eighty-five articulated and twenty-nine disarticulated skeletons were recovered. They included 65 adults (23 males, 41 females, 1 unsexed) and 26 children. Additional bone recovered from grave fills and other features brings the total minimum number of individuals to 106. The bodies were all supine and some graves provided evidence for wooden coffins. Very few grave-finds were associated to this Late Saxon cemetery, only three brooches and a couple of knives. Of the 85 skeletal remains excavated, 59 dental samples were used for the DNA analyses. Layout of the cemetery and location of their graves is shown in Figure II.10. Table II.5 shows details of sex, age and other characteristics of the individuals.

Note – hereinafter this site will be referred to as 'Norwich'.

Grave	Sex	Age	C-Date	DNA
area 1				
11305	male	old		1
11011	male	young		2
10320	n/a	18m		3
11517	n/a	n/a		4
11521	n/a	4-5		5
11535	male?	16-18m		6
11560	female	mid age		7
11570	male?	mid age-old		8
11574	female	young		9
11577	female	mid age-old		50
11578	female	mid age		10
11612	n/a	18m		11
11657	male	mid age-old		12
11622	n/a	3-4		13
11650	n/a	12m		14
11653	female	21-25	930-1020	15
11654	n/a	18m		51
11659	female	mid age		16
11665	female	mid age		52
11669	female	young	950-1020	17
11681	n/a	6-7		18
11689	female	mid age	920-1020	19
11701	male	mid age-old		20
11723	female	mid age-old	960-1020	53
11761	n/a	3-4		54
11764	n/a	4-5		55
11770	male?	mid age		56
11773	n/a	3-4		57
11778	n/a	6m		58
11784	n/a	7-8		59
11786	female	mid age?		21
11788	n/a	2-3		22
11810	female	mid age?		23
60332	?			24
60381	female	mid age-old	930-1020	25
60415	male?	mid age		26
60424				27
60433	female?	young		28
60437	female	old		29
60444	female	mid age		30
60458	female	mid age-old	960-1030	31
60466	female	mid age-old	940-1030	32
60501	?	15-18		33
60525	female	mid age		34
60528	female	mid age?		35
60541	male	mid age-old	970-1050	36
60545	female?	young		37
60560	n/a	6-8		38
60572	female	mid age		39
60582	female?	young		40
60590	female	young		41
60614	female	young		42
60629	female	young		43
60632	female	young		44
area 2				
22031	female	old?	940-1020	45
22051	n/a	3-4		46
22072	female?	mid age?		47
area 45				
45089				48
45210*	n/a	c.3-4	570-980	49

Table II.5 Details of the skeletal remains used for DNA analyses from the Farmer's Avenue Castle Mall cemetery at Norwich. Age is indicated in years. (C-date): Radiocarbon dates. (DNA) corresponds to the label during the DNA analyses.

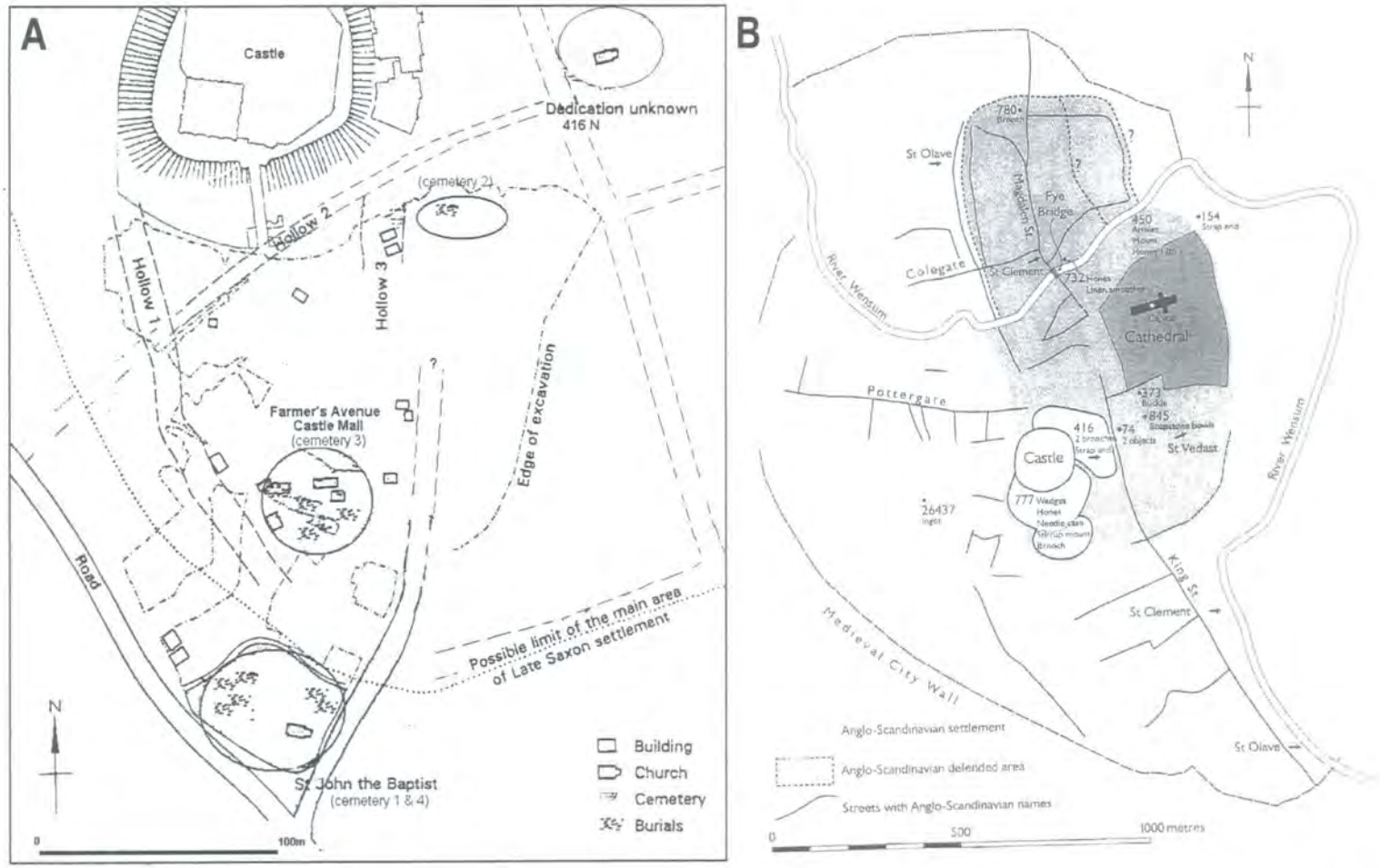


Figure II.9 Location of the Farmer's Avenue, Castle Mall site (cemetery 3) at Norwich in relation with (**A**) other cemeteries excavated at Castle Mall sites (cemeteries 1, 2 and 4), and (**B**) other Anglo-Scandinavian settlements in the area (sites 416, 74, 845, 373, 450, 732, 780, 154 and 26467). The Castle Mall site is referred to as 777. Viking artifacts are indicated.

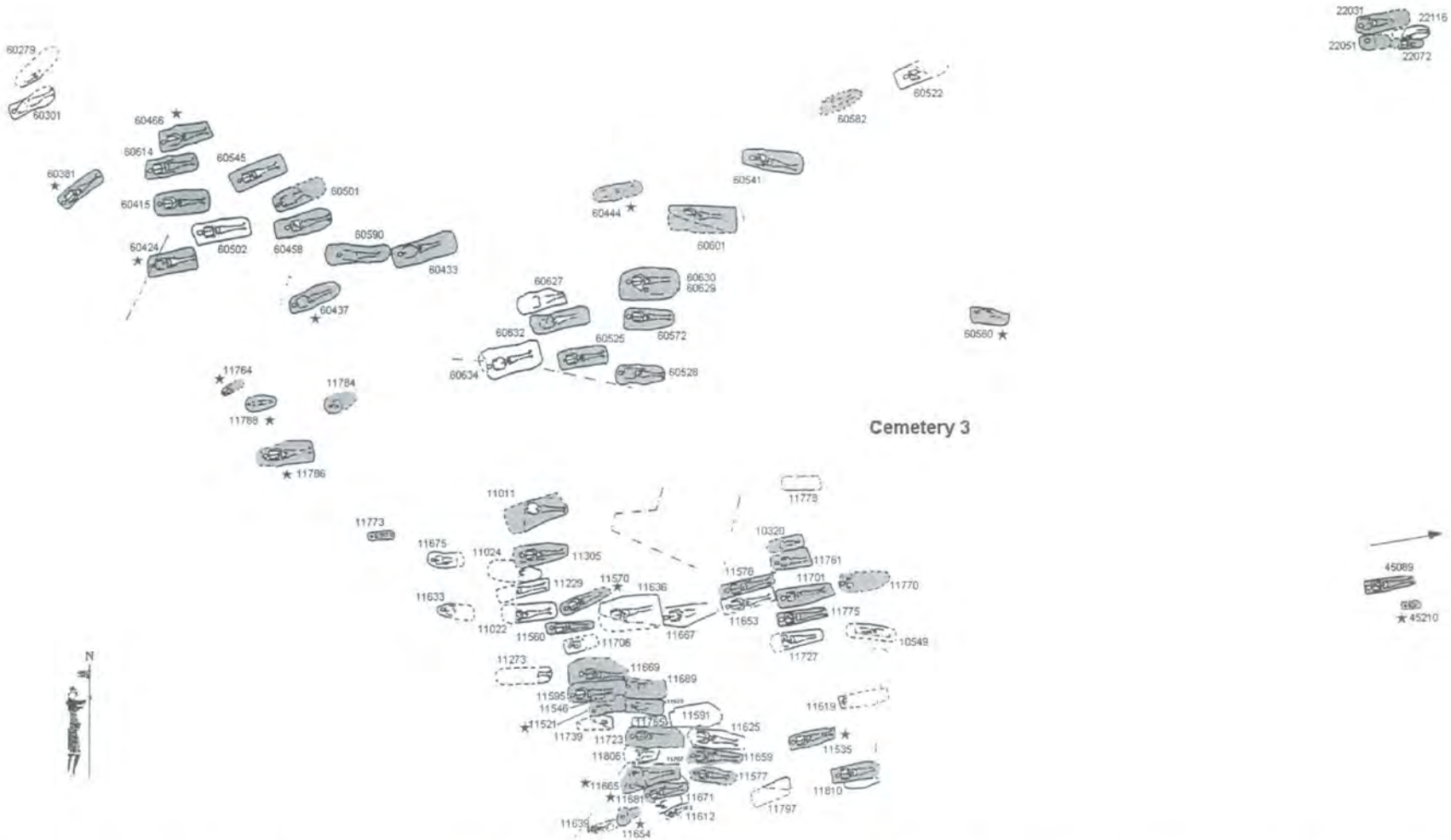


Figure II.10 Layout of the Farmer's Avenue cemetery, showing all the burials excavated (those that are further right are indicated). Location of the 59 skeletal remains used for DNA analyses is shown in grey. Samples for which DNA was obtained are further indicated with a star.

II.1.2 Modern sequence data

Mitochondrial DNA sequences from modern human populations were also analysed. An extensive database comprising 6406 mtDNA sequences from European, Asian and Middle Eastern countries was compiled for this study. DNA data was obtained from on-line public databases, such as HVR database (http://www.hvrbase.de/), NCBI (http://www.ncbi.nlm.nih.gov/), MITOMAP (http://www.gen.emory.edu/mitomap.html) and Mitocondrial DNA Concordance (http://shelob.bioanth.cam.ac.uk/mtDNA/) and personal communications or otherwise deduced from the respective publications. (Note that several sources presented errors that needed correction; see Forster 2003). Geographical location, sample size per population and original source are shown in Table II.6 and Figure II.11.

Figure II.11 Geographical location of the modern populations included in this study.

Population	Size	Original source(s)
Armenia	191	Richards *et al.*, 2000
Austria	116	Handt *et al.*, 1994 (15); Parsons *et al.*, 1998 (101)
Basques	105	Bertranpetit *et al.*,1995 (44); Corte Real *et al.*, 1996 (61)
Belarus	55	Belyaeva *et al.*, 2000
Bulgaria	141	Calafell *et al.*, 1996 (30); Richards *et al.*, 2000 (111)
Czeck Rep.	83	Richards *et al.*, 2000
England:		
Cornwall	86	Richards *et al.*, 1996 (69); Miller *et al.*, 1996 (17)
mainland	258	Helgason *et al.*, 2001 (140); Piercy *et al.*, 1993 (100)
		Anderson *et al.*, 1981(1); Miller *et al.*, 1996 (17)
Estonia	149	Richards *et al.*, 2000
Denmark:		
mainland	40	Richards *et al.*, 1996 (24); Miller *et al.*, 1996 (16)
Baltic Is.	8	Richards *et al.*, 1996
Finland	176	Kittles *et al.*, 1999 (73); Pult *et al.*, 1994 (23)
		Richards *et al.*, 1996 (30); Sajantila *et al.*, 1995 (50)
France	64	Rousselet *et al.*, 1998 (50); Miller *et al.*, 1996 (14)
Georgia	124	Reidla *et al.* ,1999
Germany:		
north	107	Richards *et al.*, 1996 (107)
central	117	Hofmann *et al.*, 1997 (67); Baasner *et al.*, 1998 (50)
south	200	Lutz *et al.*, 1999
Bavaria	49	Richards *et al.*, 1996
Greece	125	Richards *et al.*, 2000
Iceland	467	Helgason *et al.*, 2000 (394); Sajantila *et al.*, 1995 (39)
		Miller *et al.*, 1996 (20); Richards *et al.*, 1996 (14)
Ireland	101	Richards *et al.*, 2000
Italy:		
Sardinia	69	Di Rienzo *et al.*, 1991
Tuskany	49	Francalacci *et al.*, 1996
North Italy	68	Stenico *et al.*, 1996
Rome	48	Richards *et al.*, 2000
Sicily	90	Richards *et al.*, 2000
Karelia	83	Richards *et al.*, 2000
Norway	565	Opdal *et al.*, 1998 (216); Miller *et al.*, 1996 (26)
		Helgason *et al.*, 2001 (323)
Palestine	117	Richards *et al.*, 2000
Poland	37	Richards *et al.*, 2000
Portugal	54	Corte Real *et al.*, 1996
Romania	92	Richards *et al.*, 2000
Roma-Gypsy	130	Kaldma *et al.*, 1998
Russia:		
Adygei	50	Macaulay *et al.*, 1999
Uralic region	53	Bermisheva *et al.*, 2000
north Ossetia	106	Richards *et al.*, 2000
north Europe	40	Markina *et al.*, 2000
west siberia	38	Richards *et al.*, 2000
Volga-finn	32	Richards *et al.*, 2000
Saami:		
Norway	82	Delghandi *et al.*, 1998 (61); Sajantila *et al.*, 1995 (21)
Finland	69	Sajantila *et al.*, 1995
Sweeden	25	Sajantila *et al.*, 1995
Scotland:		
mainland	891	Helgason *et al.*, 2001
Western Is.	181	Helgason *et al.*, 2001
Is. Skye	49	Helgason *et al.*, 2001
Orkney	78	Helgason *et al.*, 2001
Spain:		
Andalusia	15	Corte Real *et al.*, 1996
Catalonia	15	Corte Real *et al.*, 1996
central	151	Pinto *et al.*, 1996 (18); Salas *et al.*, 1998 (92); Corte Real *et al.*, 1996 (41)
Sweeden	60	Sajantila *et al.*, 1996 (32); Kittles *et al.*, 1998 (28)
Switzerland	76	Pult *et al.*, 1994
Syria	49	Vernesi *et al.*, 2001
Turkey	72	Calafell *et al.*, 1996 (27), Comas *et al.*, 1996 (45)
Ukraine	18	Malyarchuk *et al.*, 2001
Wales	92	Richards *et al.*, 1996
Total	**6406**	

(continued)

Table II.6 Population, sample size and original source of the HVS-I sequence data included in this study.

All original sources from where the sequences were taken were population genetic studies (as opposed to pedigree studies) and clearly stated that the individuals analysed were unrelated. The only exception to this was the sample from Pult *et al.* (1994) in which 3 individuals were later reported to be maternally related [Bandelt, 2001 #867]. Two of them were therefore excluded from this study. In general, authors claimed that their samples were representative of the population analysed. Corroboration of this is beyond my possibilities. Nevertheless, for the samples from Norway, Adygei and Uralic Russia the authors provided further clarifying details (Calafell, Bermisheva, Helgason and Macaulay, pers.comm.).

As expected, the compiled raw HVS-I sequences were of different lengths. Sequences were aligned using the software SEQUENCHER 3.0 (Gene Codes Corp.) and the bases before np 16123 and beyond np 16329, were omitted. In this way, all the modern and ancient sequences would comprise the same HVS-I region (extending from 16123 to 16239) and could be used for comparative analyses. As opposed to other studies, deletions and/or insertions were not omitted.

II.1.3 General reagents

Solvents, chemicals and commercial kits were obtained from Sigma Chemical Corp., Perkin-Elmer Corp., Promega Corp., BDH Life Science and QIAgen Ltd. All reagents were used and stored according to the manufacturer's recommendation, unless stated otherwise. Buffers and solutions are detailed in Appendix A.

Enzymes and Oligonucleotides – enzymes used, namely *Taq* Gold DNA polymerase and Proteinase K, were obtained from Perkin Elmer and Sigma, respectively. Oligonucleotides were syntetised by Cruachem Ltd. and LifeBiotech.

Plasticware – falcon and eppendorf tubes, spin columns, PCR plates and tips were obtained from Greiner, Starstedt, Eppendorf, ART and Anachem. Plasticware was sterilised by autoclaving three times.

II.2 LABORATORY METHODS

II.2.4 ancient DNA work and contamination

II.2.4.1 DNA carry-over – spatial separation of working areas

In DNA analyses the most common source of contamination is the so-called DNA carry-over, that is, the contamination of samples with PCR products, usually by DNA aerosols (Kwok & Higuchi, 1989). To minimize the risk of this type of contamination, pre- and post- PCR analysis were carried out in spatially separate laboratories located in buildings 100 yards apart from each other. Human DNA analyses had never been undertaken in either of these rooms.

Pre-PCR methods included decomposition of the samples, DNA extraction and PCR set-up. Decomposition of samples was carried out in a lab in the Department of Biological Sciences. DNA extraction and PCR set-up were carried out in a fully equipped laboratory dedicated solely to the analysis of ancient DNA and isolated from any molecular biology research (Mountjoy Building, hereinafter the '*ancient DNA lab*'). PCR cycling and post-PCR analyses (i.e. gel electrophoresis, PCR product purification, etc.) were performed in a third lab (Lab 11, hereinafter the '*PCR lab*'). Lab work was carried out in a "one way procedure" daily basis; from the *ancient DNA lab* to the *PCR lab.* In this way, PCR products were never manipulated before working on either the DNA extraction or the PCR set-up, hence avoiding imperceptible carrying of DNA aerosols into the *ancient DNA lab* on clothes or skin (MacHugh *et al.*, 2000).

II.2.4.2 Modern DNA – protection and cleaning

In ancient samples, DNA is present in very small quantities so that even traces of modern material are likely to present larger amounts of DNA. In human aDNA analyses unlike non-human species, modern material is persistently available as contaminant. Skin cell, hair, sweat and saliva of the researchers or people handling the samples, as well as in reagents or disposable material, are a constant potential source of modern DNA. Stringent precautions and controls were taken to provide the best of limiting contamination.

Air-borne contamination – protective clothing covering full body and consisting of lab coat, hair net, plastic hood and sleeves, breathing mask and face shield was worn throughout all pre-PCR procedures. Two pairs of gloves were used, with the internal pair taped to the lab coat to prevent exposure of the wrist skin and the external one regularly changed. Dental samples were always handled with sterile disposable forceps. Decomposition of the dental samples was carried out in a clean fume hood isolated by a glass window and positive air flow. DNA extraction and PCR set-up was carried out in a laminar flow workstation (MicroFlow) located in the *ancient DNA lab.*

Surfaces and equipment – bench tops, instruments, racks and pipettes were thoroughly and regularly cleaned with diluted bleach. A specific DNA decontamination solution (DNA away™, Molecular Bioproducts Inc.) was also used. During non-working hours bench tops were irradiated with Ultra Violet (UV) light. The technique has been proposed by Sarkar and Sommer (1993) as an effective way of remove contamination from PCR reagents and other material. By UV-radiation, the remnant traces of DNA would be cross-linked and unable to serve as template for subsequent amplification in case of contamination

Plastic-ware and reagents – Rainin aerosol resistant filtered-sterile tips and Shaftgard filtered-sterile tips (Anachem Ltd.) were used to prevent aerosol and/or cross DNA contamination between samples. As contamination with traces of modern DNA has been found in disposable material and chemical reagents (Schmidt *et al.*, 1995), Falcon and Eppenforf tubes, Spin Columns and PCR tubes were autoclaved at least three times and then UV-irradiated for 30 min. Commercial solutions and reagents were aliquoted in small fractions and UV-irradiated for 30 min. Sterile PCR water was used for the amplification reactions. During the PCR set-up, the PCR master mix containing all the reagents except the target DNA and the enzyme were exposed to UV for 20 min, with occasional mixing. The enzyme was added, mixed thoroughly and distributed into pre-UV-irradiated PCR tubes. Only after this, the aDNA solution was added to the reactions (Kwok & Higuchi, 1989). Ready PCR reactions were brought to the *PCR lab.* The positive PCR control (high molecular weight DNA) was set up just before starting the PCR reaction (i.e when the ancient samples were no longer manipulated). After amplification, PCR products were kept in the *PCR lab.*

II.2.4.3 Monitoring of contamination

DNA extraction controls – DNA extractions were carried out in groups of 12 samples; 10 ancient samples plus 2 controls carried alongside the whole extraction procedure. These two latter were: a) an extraction control, containing all the extraction reagents but no material source of DNA (i.e. dental powder), which allowed checking out for contamination in the extraction reagents, and b) a carrier control, which contained all the extraction reagents and an aliquot of a low concentrated non-human (plant) DNA. It has been shown that the presence of an extra source of DNA may evidence other contaminating DNAs that would not be detected otherwise (Kolman & Tuross, 2000). Both controls were subjected to PCR amplification along with the rest of the samples.

PCR control – to identify contamination in the PCR reagents, a PCR blank containing all the standard PCR reagents but water instead of a DNA solution was included with each PCR amplification reaction lot.

DNA sequences – if extraction and/or PCR controls happened to be contaminated they were sequenced so that the contaminating sequence was determined. As one of the most likely sources of contamination is believed to be from the researches themselves, mtDNA sequences from individuals working in the *ancient DNA lab* and adjacent areas, namely AT, RH, RF, JT and TG, were obtained and recorded for comparisons

II.2.5 Cleaning and decomposition of samples

External dirt (e.g. soil, clay, debris or bone debris) adhered to the dental samples was cleaned off with a hard brush and distilled water. Teeth were then soaked with constant agitation in 3% hydrogen peroxide or 30% bleach, rinsed in ultra pure distilled water and soaked in 70% ethanol for 30 min. each (Ginther *et al.*, 1992). Then, the ethanol was discarded and the teeth air-dried. To eliminate any further possible external contamination, samples were UV-irradiated (Sarkar & Sommer, 1993). Teeth were placed under a 254 nm UV lamp and each face of the tooth exposed for approximately 20 min.

Samples were wrapped in sterile plastic metal foil and open lengthwise by pressure with a small mechanical vice. If loosen, pieces of enamel were discarded so that only the internal part of the tooth was utilized. Dental pulp or otherwise its remnants were

obtained by grinding the walls of the pulp chambre with an electric drill (Dremmel). To facilitate this procedure the (cordless) drill was fixed with a clamp to a retort stand (so that the drill did not need to be handled). The dremmel was switched on and the dental samples pressed against the spinning dremmel bit. The fine powder produced was collected in a sterile tube placed below the sample. Usually, 50 to 70 mg of dental powder were collected. This procedure would take between 5 and 15 minutes per sample depending on the condition and hardness of the tooth. The remaining pieces of tooth, excluding the enamel, were wrapped in foil and completely pulverized with the vice. The resulting powder and small pieces were kept for replicate analyses. Between decomposition of different dental samples, the dremmel bits were soaked in diluted bleach, rinsed in distilled water and UV-irradiated for 20 min.

A total of 319 dental samples were used (62 from Buckland, 36 from Leicester, 56 from Lavington, 47 from Norton and 118 from Norwich) and a total of 721 fractions of powdered dental material were subjected to the DNA extraction.

II.2.6 DNA isolation

Several DNA extraction protocols were tried out. As will be shown in Chapter III, the method described below yielded the best results. It is an extremely sensitive method, developed for this work based on a protocol previously described by Schmerer et al. (1999). It consists of an enzymatic digestion followed by an organic extraction and a silica-based concentration.

II.2.6.1 Enzymatic digestion

Powdered dental samples were incubated with 1000 µl of Lysis Buffer (0.5% SDS, 0.45 M EDTA pH8, 10 mM Tris-HCl) (Montiel, pers.comm.) and 20 µl of Proteinase K (50 mg/ml) with constant agitation in a rotary wheel. Fine-powdered samples were incubated at 56 °C overnight, whereas samples containing bigger pieces of material were incubated at 56 °C for 48 to 60 hours. In this latter case, 10 µl of Proteinase K were added half way through the incubation. After this, tubes were centrifuged at maximum speed (13,000g) for 3 to 5 min. to spin down the remnant material. For the fine-powdered samples this was not necessary as after the incubation they were

completely disintegrated. Aliquots of 500 µl of the supernatants (containing the lysate cells from the dental powder) were subjected to the organic extraction. The rest of each sample was stored at -4 °C.

II.2.6.2 Organic extraction

Tris-saturated phenol was used to denature and subsequently remove protein impurities from the aqueous DNA solutions. 500 µl of each supernatant was transferred to a sterile 1.5 ml eppendorf and an equal volume of phenol pH 7.6 added. The mix was vortexed for 1 min. and centrifuged for 3 min. to allow phases to separate. Due to the high concentration of EDTA in the lysis solution, the aqueous and organic phases were inverted. The upper organic phase was carefully discarded and the bottom aqueous phase transferred to a clean tube. An equal volume of phenol:chloroform:isoamilalcohol 25:24:1 (v/v) was added, vortex mixed and centrifuged as before. The aqueous phase (now the upper one) was transferred to a new clean tube, taking care not to dislodge the protein interface between the two layers. This extraction process was repeated once again with an equal volume of chloroform:isoamilialchol 24:1 (v/v). Chloroform helps to remove lipids from the aqueous layer and isoamyl alcohol is used as an anti-foaming agent. The final clean aqueous phase was transferred to a clean eppendorf tube for the concentration step. For this step, commercial sterile and individually wrapped tubes (Eppendorf) were used to minimize possible contamination.

II.2.6.3 Silica based concentration

To concentrate the aqueous DNA solution, a silica-based method was used. In presence of a caotropic agent, DNA binds to silica particles (Hoss & Pääbo, 1993). The DNA-silica binding is pH dependent, with pH ≥6 being the optimum (Schmerer *et al.*, 1999) (QIAEXII kit manual, 2000). The commercial caotropic solution QX1 Buffer (QIAgen) contains a pH indicator which turns from yellow (to orange) to purple as its pH increases. This buffer was preferred for the concentration step so that the pH of the solution could be easily controlled.

Following the instructions given by the manufacturers in the manual, three volumes (approx. 1.5 ml) of QX1 Buffer and 10 µl of the silica suspension (QIAgen) were added

to the aqueous DNA solutions. Samples of basic pH (indicated by the buffer turning to orange or purple) were equilibrated to the appropriate pH with 5 to 10 µl of 3 M AcNa, pH 5.0, until the colour of the solution turned back to yellow. Samples were incubated under slow agitation at 60 °C (Smith *et al.*, 1995) for 15 to 20 min. to allow the formation of the DNA-matrix complexes. After this, the samples were filter through a spin filter column (Anachem) to eliminate the caotropic agent (i.e. QX1 Buffer). Approximately 600 µl of the silica-DNA-QX1 Buffer solution were loaded into the column and centrifuged for 45 sec. The complexes of DNA-silica were retained onto the filter and the caotropic agent filtered through and discarded. This was repeated until the total volume (approx. 2 ml) was filtered through the column (one single column per sample was used). The DNA-silica matrixes retained on the filter were washed with 500 µl PE Buffer (ethanol-based buffer, QIAgen). Samples were centrifuged at maximum speed for 45 sec. and the buffer discarded. This wash was repeated twice. The sample pellets were centrifuged for 120 sec. so they dried out completely, as any traces of ethanol residue may inhibit subsequent enzymatic reactions (QIAgen manual). As in presence of a basic solution the DNA-silica complexes separate, the DNA was eluted from the filters in 50 µl of EB Buffer (TE-based buffer, QIAgen). Samples were centrifuged at maximum speed for 30 sec. and the DNA solution was collected in a clean tube. DNA solution was kept at 4 °C for 3 days and then stored at -20 °C (Montiel *et al.*, 1997), until being used for the PCR reaction.

II.2.7 PCR amplification

A 264-bp fragment of the hypervariable region I (HVS-I) of the mitochondrial DNA control region was amplified by PCR. Primers for the control region (whole length) were originally described by Vigilant (1989). However, this set of primers flank a 541 bp fragment, which was impossible to amplify from aDNA (see Chapter III.2.3.1) New primers were specially designed so that the target region was of a length feasible to be amplified from aDNA. Primers sequences were:

5' – AACCGCTATGTATTTCGTAC – 3'

5' – TTTGACTGTAATGTGCTATGTA – 3'

and amplified fragment of 264 bp from positions 16099 to 16331 of the mtDNA HVS-I.

Each 25-µl PCR reaction contained 0.3 µM each primer, 0.2 mM each dNTP, GeneAmp® 1x PCR Gold Buffer (15 mM Tris-HCl (pH 8.0) and 50 mM KCl, Perkin Elmer), 15 mM $MgCl_2$, 0.08 µg/µl BSA and 1U *Taq* Gold™ DNA polymerase (Perkin Elmer). As mentioned, the PCR master mix was first UV-irradiated for 20 min., then the enzyme added and the mix distributed in aliquots of 20 µl into UV-irradiated PCR tubes. Finally, 5 µl of the aDNA final extraction solution were added to the PCR reactions and mixed carefully (Kwok & Higuchi, 1989). During the PCR set-up samples were kept in ice even though the polymerase enzyme remains inactive at room temperature. PCR reactions were overlaid with a drop of mineral oil to avoid evaporation of the solutions during the high temperature steps of the PCR cycling. Samples were then brought into the *PCR lab*. The positive PCR control (high molecular weight DNA) was set up just before starting the PCR reaction (i.e when the ancient samples were no longer manipulated). The amplification was carried out in a Perkin Elmer TC1 480 thermal cycler located in the *PCR lab*. Cycling conditions were as follows:

Stage 1 — 94 °C for 11 min. (enzyme activation)

Stage 2 — 94 °C for 45 sec. (denaturation)

53 °C for 45 sec. (annealing)

72 °C for 90 sec. (elongation)

Stage 3 — 72°C for 5 min. (final extension)

Following Rameckers et al. (1997) the amplification cycle (i.e. stage 2) was carried out for 45 cycles, to allow the amplification of the few template DNA strands. Once the programme was finished the reactions were cooled down to 4 °C and stored in the *PCR lab* until being run in an electrophoresis gel.

II.2.8 PCR products - gel electrophoresis and purification

Analyses of amplification reactions were carried out by agarose gel electrophoresis. Gels consisted of 1.5% SeaKem LE agarose gel (FMC Corp.) that efficiently resolves short PCR products. The agarose was mixed with the required volume of 1x TBE Buffer (see Appendix A) and dissolved by microwaving. After cooling down, the agarose was poured into a gel mould with the comb in place. Once set, the gel was placed in a Pharmacia Biotech horizontal gel apparatus (GNA-100 electrophoresis tank) and

covered with 0.5x TBE. Eight-µl of each PCR product was mixed with the appropriate volume of 6x Loading Buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF and 40% sucrose, in distilled water) and loaded into the wells. Two-µl of 100 bp DNA Ladder (MβI Fermentas) was loaded alongside. Running conditions were 100 V for 45 min. The gel was stained in 100 ml of water containing 1 µl of ethidium bromide (10 mg/ml) for 10 min and distained for another 20 min in clean water. Bands were visualised by UV transillumination and photographs taken to keep record of the results.

Amplification products showing a band of the right size were purified using QIAquick PCR purification kit (QIAgen). For this, 5 volumes of PB Buffer were added to the PCR reactions and mixed thoroughly. Samples were loaded in a spin column and centrifuged for 45 sec. at high speed. PCR products retained onto the filter were washed with 300 µl of PE Buffer (ethanol-based buffer). The buffer was removed by centrifugation for 30 sec. and the wash was repeated. The samples were further centrifuged for 2 min. to ensure that no trace of ethanol was present, as this may inhibit the sequencing reaction. PCR products were recovered by adding 30 µl of TE Buffer to the centre of the filter, leaving to elute for 2 min. and centrifuging for 45 sec. PCR fragments were then ready for the sequencing reaction.

II.2.9 DNA Sequencing and alignment

PCR fragments were sequenced directly with the ABI BigDye Terminator sequencing kit™ (PE, Applied Biosystems). Sequencing reactions consisted of 1 µl primer (3.2 pm/µl), 8 µl sequencing mix, 5 to 10 µl purified PCR product and distilled water up to 20 µl. The standard cycling conditions were: 96 °C for 30 sec., 50 °C for 20 sec. and 60 °C for 4 min, for 25 cycles. Sequencing products where purified using the QIAquick PCR purification kit (QIAgen) as described above and then run on a 6% denaturing polyacrylamide DNA sequencing gel on an ABI PRISM 377 (or 373) automatic DNA sequencer. At the end of the run an output file (i.e. DNA electropherogram) was produced for each DNA sequence.

In principle, one strand, namely the light strand (L) was sequenced, as mtDNA sequencing of the heavy strand (H) has proven to yield unclear sequences due to the high purine (A and G) content (Parson *et al.*, 1998). For this, the same forward primer as for the PCR reaction (L16099) was used. When a blurred sequence was obtained,

usually due to the presence of an uninterrupted polycytosine tract around the nucleotide position 16189 (Bendall & Sykes, 1995), PCR products were re-sequenced using an internal primer (L16179), resulting in a 195 bp fragment. This primer anneals at the cytosine tract and thus allows the clear reading of the sequence beyond this point. Moreover, to further corroborate such samples, these PCR products were also sequenced for the reverse strand using primer H16331.

For all the sequences, the first tens of nucleotide positions (np) were discarded as they were usually unclear (due to the overlapping of very short sequences). The ending 20 base pairs were also discarded as were the reverse primer sequence. Thus, the final DNA sequences were of 207 bp, from positions 16123 to 16329, numbered according to the CRS (Anderson *et al.*, 1981). All internal dubious positions caused by errors during the reading of the ABI software or to sequence heteroplasmy were carefully checked by hand by comparisons with replicate sequences. Once sequences were confirmed, corrected output files were loaded into the software Sequencher 3.0 (Gene Code Corp., Ann Arbor MI, USA) for alignments. Variable nucleotide positions were identified in reference to the CRS and recorded for posterior analyses.

II.3 Analytical Methods

II.3.1 Data analysed – subdivisions and groupings

All calculations were based on the 207-bp HVS-I sequences from the five settlements and 50 modern populations. For some analyses, the ancient communities were also analysed individually as well as in groups according to their historical period. The groups were defined and composed as follows:

— *Romano-British* period (300-400 AD): Newarke Street cemetery at Leicester.

— *Early Saxon* period (450-750 AD): Norton at Cleveland, Grove Farm, Market Lavington at Salisbury and Dover, Buckland at Kent.

— *Late Saxon* period (850-1050 AD): Farmers' Avenue, Castle Mall at Norwich.

In addition, for comparative analyses a mixed sample was artificially composed by pooling together the modern populations which would be expected to represent those ancient populations that might have been admixed in the ancient communities.

— *Britons:* England, Cornwall and Wales

— *Saxons:* north Germany and Denmark

— Vikings: Norway

II.3.2 Intra-population analyses

II.3.2.1 Haplotype diversity

Several indexes were used to measure the haplotype diversity within populations. Number of segregating sites, transition and transversion rates were also calculated.

Proportion of haplotypes – was calculated simply as *k/n*, where *k* is the number of observed haplotypes and *n* is the sample size.

Gene diversity – $\hat{h}$ is the equivalent of the expected heterozygosity for diploid data and represents the probability that two randomly chosen sequences from the sample are non identical by state. Gene diversity was estimated following Nei, (1987), p.180, as

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

where n is the number of gene copies or DNA sequences in the sample, k is the number of haplotypes and p_i is the sample frequency of the i-th haplotype. Gene diversity was calculated with the software package ARLEQUIN 2.1 (Schneider *et al.*, 2001). To avoid confusion with the Shannon index (see below), gene diversity will be referred to as *GD* (instead of $\hat{h}$).

Shannon index – H_S is much used in Ecology to characterize species diversity in a community. It accounts for both abundance and evenness of the species (in this case mtDNA sequences) present and it is defined as:

$$H_S = -\sum_{i=1}^{s} p_i \ln p_i$$

where p_i is the proportion of haplotypes i relative to the total number of haplotypes.

As this index is strongly dependent on the sample size, a standardized (corrected) version was used instead (H'_S) (Magurran, 1988). This is simply the division of the H_S calculated by the highest possible value of H_S for a given sample size. That is,

$$H'_S = \frac{H_S}{-\ln(1/n)}$$

where n is the sample size. A value H'_S=1 would imply that all the individuals in a sample bore different sequences. This index was calculated by hand.

Allelic richness – denoted as r(g) and originally defined for biallelic systems is the number of different alleles found when g genes –the specified sample size– are sampled. This index is corrected for sample size by the rarefaction method. This method allows comparing the number of haplotypes found in two regions when the sampling effort differed. The method uses the data from the larger sample to determine how many haplotypes would have been found in a smaller sample. Thus, if a total of N ($N > g$) individuals are analysed in the larger sample, the expected number of different haplotypes in a sample of g individuals can be obtained by the formula

$$\hat{r}(g) = \sum_i \left[1 - \frac{\binom{N - N_i}{g}}{\binom{N}{g}} \right]$$

where N_i represents the number of occurrences of the i-th allele among the N sampled genes. No underlying distribution of allelic frequencies is assumed to derive $\hat{r}(g)$; the formula is purely based on the observed relative abundance of the alleles in the sample of N genes. The rarefaction method was proposed by Sanders (1968) and corrected by Hurlbert (1971) and Simberloff (1972). The corrected version was calculated with the software RAREFACTOR calculator (Brzustowski, 2001). The programme also gives the standard deviation of the index.

II.3.2.2 Nucleotide diversity

The mean numbers of pairwise differences (π) between all pairs of sequences in the sample was calculated following (Tajima, 1983) as

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

where $\hat{d}_{ij}$ is an estimate of the number of mutations having occurred since the divergence of haplotypes *i* and *j*, *k* is the number of distinct haplotypes and p_i and p_j are the respective frequencies of haplotypes *i* and *j*.

There are many methods for estimating the evolutionary distance $\hat{d}_{ij}$ depending on the pattern of nucleotide substitution. The Tamura-Nei method was preferred for the calculations, as it assumes higher rate of transitional than transversional substitutions and different rates between purines (A and G) and pyrimidines (C and T) transitions. To account for the heterogeneous distribution of the polymorphic sites along the mtDNA CR a gamma distribution with $\alpha = 0.26$ (Meyer *et al.*, 1999) was applied. This value seem to be more accurate than the traditional $\alpha = 0.47$ (Wakeley, 1993). The total variance (over the stochastic and the sampling process) assuming no recombination between sites and selective neutrality, was obtained following Tajima (1983).

II.3.2.3 Mismatch distributions and expansion dates

According to Rogers and Harpending (1992), the distribution of pairwise differences reflects the demographic past of the population. Thus, for a demographic stable population this distribution is multi-modal whereas in populations under expansion or decline the distribution is unimodal. The distribution of the observed pairwise nucleotide site differences (i.e. the mismatch distribution) for the populations and the expected curves under equilibrium or expansion were obtained by means of DNASP 3.5.3 (Rozas & Rozas, 1999). The programme gives the three parameters on which the model is based: θ_0 (theta initial, i.e. theta before the population growth or decline), θ_1 (theta final, i.e. theta after the population growth or decline) and τ, the date of the growth or decline measured in units of mutational time ($\tau = 2ut$) by generating 10,000 random samples. The program also computes the raggedness index (*r*) (Harpending, 1994; equation 1) which quantifies the smoothness of the observed pairwise differences. Larger values of *r* are usually obtained for multimodal distributions (typically found in stationary populations) than for unimodal distributions (typically found in expanding populations).

The modern and ancient population of England were also analysed for Tajima's test for selective neutrality, which is based in the infinite-site model without recombination (Tajima, 1989) in DNASP 5.3.3 by generating 10,000 random samples. Significant *D* values are theoretically interpreted as deviation from neutrality, however, it should be noted that also other factors may produce this result. Population expansions and bottlenecks, for example, would also result in *D* values significantly different from zero. After a bottleneck, rare mutations are lost more easily than the common ones, resulting in a momentary positive *D* values. On the other hand, for populations in expansion, an increase in population size will result in a temporary excess of new rare mutations which will result in negative *D* values. Thus, the *D* values were used as another way of studying the demographic state of the populations.

For some of the populations, the expansion dates were calculated as

$$t = \frac{\tau}{2u}$$

where t is the generation time, τ is the expansion time measured in units of mutational time and u is the mutation rate of the whole sequence (i.e. $u = \mu L$, with L the length of the sequence analysed and μ the mutation rate per site per million years). In this case, L = 207 bp. Given the important disagreement regarding the mutation rate of the CR, different rates were used for the calculations: $8.65\ 10^{-8}$ mutation/site/Myr (Vigilant, *et al.*, 1991); $1.03\ 10^{-7}$ mutation/site/Myr (Horai *et al.*, 1995); $1.8\ 10^{-7}$ mutation/site/Myr (Forster *et al.*, 1996) and $4.5\ 10^{-7}$ mutation/site/Myr (Howell *et al.*, 2003). This latter is a pedigree-derived estimation of the mutation rate whereas the other three are phylogenetically derived.

II.3.2.4 Population size

Several methods were used to estimate the population parameter $\theta = 2N_{fe}\mu$, where N_{fe} represents the female effective-population size and μ is the mutation rate. Because the mtDNA control region mutation rate should be the same in all populations, differences in θ values reflect differences in N_{fe} –that is the mean number of women who have transmitted their mitochondrial DNA to female offspring during past generations.

Theta π (θ_π) – is based on the observed nucleotide diversity and therefore estimates the female effective population size that would allow such number of pairwise differences between the sequences to arise through mutation events in a single population. This means that it tends to reflect the mean of N_{fe} over long periods of time. Theta π was calculated as above described for the mean number of pairwise differences (π) (see section II.10.2).

Theta S (θ_S) – is based on the number of polymorphic sites assuming an infinite site model of evolution. It is obvious that the number of segregating sites depends on the number of DNA sequences used. As the number of sequences increases, the expectation of the number of segregating sites also increases. To convert the number of segregating sites into an estimate of genetic diversity that does not depend on the number of sequences, it must be assumed that the mutations are selectively neutral and that the population is panmictic and at equilibrium. Watterson (1975) showed that the expectation of the number of segregating sites is

$$\hat{\theta} = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}}$$

where S is the total number of segregating sites in the sample and n is the number of sequences.

Theta k (θ_k) – is based on the number of haplotypes, assuming the infinite-allele model of molecular evolution and was estimated following Ewens formula (1972)

$$E(k) = \theta_k \left(1 + \frac{1}{\theta+1} + \frac{1}{\theta+2} + \cdots + \frac{1}{\theta_k + n + 1} \right) \text{ that is, } E(k) = \theta_k \sum_{i=0}^{n-1} \frac{1}{\theta_k + i}$$

where k is the number of distinct lineages observed in a sample size of n. k is also dependent on the sample size and assumes that mutations are neutral and that the population is at equilibrium.

According to Helgason *et al.* (2000) the estimators θ_S and θ_k based, as they are, on the relationship between sample size and the number of polymorphic sites, and between sample size and the number of distinct lineages, respectively, are more sensitive to the effects of lineage sorting during recent demographic history.

The three estimators and their standard deviations (SDs) or 95% confidence interval (CIs), were calculated with ARLEQUIN 2.1 software (Schneider *et al.*, 2001) for the modern and ancient populations. The five ancient communities were also analysed individually, as well as grouped according to their historical period as described earlier.

II.3.3 Genetic differentiation between populations

II.3.3.1 F_{ST} statistics

Genetic differentiation between populations was quantified by F_{ST} statistics. Population pairwise estimators were calculated in ARLEQUIN 2.1 based on sequence data. Permutation for significance was set to 10,000 and *P*-values represent the proportion of permutations showing an estimator larger or equal to the observed one. For multiple comparisons, as the Bonferroni correction is known to be very conservative, the sequential method of Holm (1979) (also known as sequential Bonferroni) was applied with Multiplicity 2.0 (Brown & Russell, 1996).

F_{ST} statistics were calculated for comparisons between the modern and ancient populations and also among the five archaeological settlements themselves. Given the low values obtained, an average F_{ST} was computed for each population. That is, for population X, all F_{ST} statistics between population X and population *i* (with *i* different from X) were averaged.

II.3.3.2 Nei distances

Genetic distances between populations were calculated following Nei (1987), (equations 10.20 and 10.21) as

$$d_{XY} = \sum_{j}^{i} x_i y_i d_{ij}$$

$$d_A = d_{XY} - \frac{d_X + d_Y}{2}$$

where x_i and y_i are the frequencies of the *i*-th haplotype for population X and Y, respectively, d_{ij} is the number of nucleotide substitutions per site between the *i*-th

haplotypes from X and the *j*-th haplotypes from Y and d_X (or d_Y) is the pairwise mean within population X (or population Y). As before, the Tamura-Nei method with α = 0.26 (Meyer *et al.*, 1999) and a transition:transversion ratio of 15:1 (Tamura & Nei, 1993) was assumed for the estimation of $\hat{d}_{ij}$.

d_{XY} represents the average number of nucleotide substitutions between DNA haplotypes from populations X and Y whereas d_A represents the number of net nucleotide substitution between the two populations (i.e. total divergence minus the within-population component). Therefore this distance estimates the divergence occurred *after* the two populations split. Both distances were calculated using the software package ARLEQUIN 2.1. Since the obtained distances were very small, values were multiplied by 100 for clarification. Permutation for significance was set to 10,000 and *P*-values represented the proportion of permutations showing an estimator larger or equal to the observed one after the sequential Bonferroni was applied.

Given the low genetic distance between populations, an average distance was computed for each population. That is, for population X, all d_{XY} (or d_A) distances between population X and population *i* (with *i* different from X) were averaged. They were referred to as $\overline{d}_{XY}$ and $\overline{d}_A$. Significance of the differences between average distance values (i.e. $\overline{d}_X$ and $\overline{d}_A$) for the different populations was evaluated by a Student t-test with α = 0.05.

II.3.3.3 Chord distances

Geometric genetic distances (i.e with no biological assumption) between populations were also calculated as chord distances (*f*) based on frequencies of haplogroups (see below). Chord distances were introduced by Cavalli-Sforza (1967) as

$$D_{C-Sf} = \sqrt{2 - 2\cos(\varphi)}$$

with $\cos\varphi = \sum_{i=1}^{k} \sqrt{S1_i S2_i}$

where $S1_i$ (or $S2_i$) or is the frequency of the haplogroup *i*-th in population 1 (or 2).

Haplogroup assignment was carried out according to the following algorithm:

hg	segregating sites				
J	126				
J1 [a]	↓	261			
J1a	↓	↓	145	231	
J1b	↓	↓	↓	222	
J1b1	↓	↓	↓	↓	172
J2	↓	193			
T	↓	294			
T1	↓	↓	163	186	189
T2	↓	↓	304		
T3	↓	↓	292		
T4	↓	↓	324		
T5	↓	↓	153		
K	224	311			
K1	↓	↓	320		
U1	249 (and 189 or 327)				
U2	129c				
U5	270				
U5a	↓	192			
U5a1	↓	↓	256		
U5a1a	↓	256			
U5b	↓	189			
U5b1	↓	↓	144		
U6	172	219			
U6a	↓	↓	278		
U6b	↓	↓	311		
U7	318t				
V	298				
IXW	223				
I [b]	↓	129			
W	↓	292			
X	↓	189	278		
A	↓	290	319		
C	↓	298	327		
Z	↓	185	224	260	298
N'	↓	147t/248 or 145a/176g			
L'	↓	278 (and 189/311)			
H'	none of the above				

Table II.7 Algorithm used for the haplogroup assignment, based on Macaulay et al. (1999). Nucleotide positions are according to the CRS minus 16000 for brevity. Arrows indicate the same segregating site as above. (a) hg J is defined by sites 126 *and* 69, however, this latter site is not available in this study; (b) hg I is defined by sites 223, 129 *and* 391, this latter is not available either.

It has to be noted that, since only a 207 bp-fragment of HVS-I was analysed not all the relevant sites for haplogroup classification of all clusters are available. For example, haplogroup J (part of super-haplogroup JT defined by 126) is defined by segregating sites 126 *and* 69, of which this latter is not available. However, as proposed by Torroni *et al.* (2000) haplogroup J can still be obtained as JT *minus* T, that is all sequences that are JT (i.e. defined by 126) and are not T (defined by 126 and 294). On the other hand, haplogroup U3 and U4 defined by HVS-I segregating sites 343 and 356, respectively, cannot be determined and will be included within H'. Neither can haplogroup H be

properly classified, as site 73 of HVS-II is not available. Thus, H' is not actually a haplogroup in the strict phylogenetic sense, i.e. it is not a monophyletic clade of the mtDNA genealogy (Torroni *et al.*, 2000) but a polyphiletic clade.

II.3.3.4 MDS graphs

To visualise any possible geographic pattering among the populations, both the Nei's distances and the chord distances were graphically represented by means of a Multidimensional Scaling (MDS) plot. MDS attempts to find the structure in a set of distance measures between objects or cases. This is accomplished by assigning observations to specific locations in a conceptual space (usually two- or three-dimensional) such that the distances between points in the space match the given dissimilarities as closely as possible. In this case the dissimilarities are the distances between populations given in a distance matrix format. Thus, the analysis resumes the 50 dimensions of the distance matrix in a 2-dimensional space. Unlike Principal Component analyses, Muldimensional Scaling procedure is relatively free of distributional assumptions. The analyses was carried out with the statistics software SPSS 11.0 set for 20 random multiple starts, a minimum stress convergence of 0.0001 and 1,000 maximum iterations.

II.3.4 Founder analyses

Rho (ρ) distances were originally defined as the average number of sites differing between a set of sequences and a specified common ancestor –which need not be among the sampled sequences (Forster *et al.*, 1996). For this, ρ is based on the topology of a rooted tree, where the common ancestor (or founder sequence) of a haplogroup can be identified. To compute ρ the number of transitions is counted from each sampled sequence to the corresponding founder sequence and then groupwise averages are taken. However, Helgason et al. (2001) made a slightly different use of ρ to estimate the mutational divergence (i.e. ρ distance) between populations involved in founder events. It assumes that after a founder event the haplotypes observed in the new population (referred to as the *founder* population) will be a random sub-sample of haplotypes from the population from where the migration originated (referred to as the *source*

population). Therefore, in the source population one should find identical –or very closely related– sequences to those observed in the *founder* population; these are the *founder* haplotypes. Thus, the index is calculated as the average number of mutational differences between the sequences of the *founder* population and the closest *founder* sequences observed in the *source* population. It effectively summarises the overlap between one mtDNA pool and a potential source mtDNA pool (Helgason *et al.*, 2001). In addition, common haplotypes shared among the majority of the populations will not account for the distance between them. This is particularly important when analysing shallow diversity as is the case for mtDNA where the divergence between populations is small relative to the overall mutational divergence of the European mtDNA.

Rho is defined as

$$\rho = \frac{1}{n}\sum_{i=1}^{n} r_i$$

where r is the distance (as mutational differences) from the i-th haplotype of the *founder* population to the closest founder haplotype in the putative *source* population and n is the size of the founder population.

To compute ρ, a distance matrix (where distances are the net number of observed mutational differences between haplotypes) between the *founder* population and a putative *source* population was obtained using the software package MEGA 2.0 (Kumar *et al.*, 1993). Then, based on this distance matrix, for each haplotype in the *founder* population, the founder haplotype in the putative *source* population was defined as the haplotype with the fewest mutational differences. Rho was finally calculated as the average value for all these distances.

As ρ distances proved to be strongly correlated with the sampling effort (i.e. sample size) of the putative *source* population a correction for sample size was applied manually. For this, each source population was randomly divided in several sub-samples of the same size. The sub-sample size was set to $n = 50$ to match the size of the ancient population –however, this does not need to be a requisite. A distance matrix (of net mutational differences) between the *entire* founder population and each *sub-sample* of the source populations was obtained. Partial ρ (ρ_{50}) distances were computed for each of these matrixes, as above described. Finally, the ρ distance between the founder population and the original source population was computed as the average of the ρ_{50}.

As an attempt to identify the origin of the ancient mtDNAs, ρ distances were computed between the ancient population of England (as founder population) and several putative source populations. In order to determine whether the communities dateing from different historical periods would reflect any differences in the alleged immigrant population, ρ distances were also calculated for the sub-groups of the ancient population as before (i.e. Romano-British, early Saxon and late Saxon groups) as founder populations.

Moreover, to evaluate the extent to which these ancient communities contributed to the modern mtDNA pool, ρ distances were computed for the modern population of England (as founder population) and several putative source populations, including the ancient samples. In addition, two control populations (Portugal and Orkney) were also analysed as founder populations against a number of source populations.

II.3.4 Phylogenetic analyses

Almost any human mtDNA data set used for phylogenetic studies, whether restriction length polymorphism (RFLP) data or sequences from the hypervariable control region, almost invariable fail to form nested sets or clusters of haplotypes. Instead it exhibits incompatibility between the characters. The reason for this is homoplasy (parallel events or reversal) in conjunction with the low number of informative characters analysed. This means that resolution is lost as the coalescent time approaches (Hedges *et al.*, 1991).

Bandelt et al. (1995) argued that mtDNA data set are best analysed by a network approach that distinguishes between irresolvable and resolvable character conflicts, leaving a representation of plausible solutions. This approach works with median networks generated by partitioning the groups of haplotypes character by character. In this way, unmodified median networks include all the most parsimony trees and can highlight character conflicts in the form of reticulations, which can then be interpreted in terms of homoplasy, recombination or sequence error.

Reduced median joining networks were generated to infer relationships between mtDNA lineages from the ancient and modern populations of England by means of the software package NETWORK 2.0 (Röhl, 1997). The network methods are designed for non-recombining DNA haplotypes, RNA or amino acid sequences. The program can

generate both median joining (MJ) and reduced median (RM) networks and accepts either binary or multistate data as input files. The former typically are short tandem repeat (STR) data. The latter are amino acid sequences and DNA sequences which show nucleotide positions with more than two different nucleotides. Multistate data can be analysed only by the median joining (MJ) network method (which is unreliable for longer branches) and not by the more robust reduced median (RM) network method. In order to be able to use the RM method, the mtDNA multistate data was transformed into binary data by splitting variants into transitions and transversions.

Pre-processing of data – unlike RFLPs, which constitute genuine binary data (since a site is either present or absent) sites in DNA sequence data can be any one of the four nucleotides A, G, C and T. Mitochondrial DNA data was transformed into binary data by determining "presence" or "absence" of a nucleotide variant by comparison with a reference sequence (CRS; Anderson *et al.*, 1981). The vast majority of the observed changes were transitions so that these sites were easily turned into binary data. For the observed transversions, it was stipulated that they had been single events, since transversions are by at least an order of magnitude less likely than transitions. Thus, they were also readily turned into binary data. Sites where more than two variants were observed, were split into (n-1) sites, where n is the number of variants. Thus, these sites were regarded as a pair or a triplet of binary data, one of which might denote a transversional change. The thus modified haplotypes could be handled as binary data.

Identical haplotypes were pooled leaving n distinct haplotypes. The number of individuals with haplotype i is the frequency f_i. Sites that were constant among the haplotypes were removed so that all sites (k) were variable. Each site of the n distinct haplotypes splits the haplotypes into two groups according to the two states at the site. These two variants were given status 0 and 1, under the convention that a certain reference haplotypes (CRS) receives status 0 for all the characters. Thus the haplotypes were represented by 0-1 vectors of length k and weight w (see below).

Weighting scheme – to further maximize the precision of the analyses all sites between positions 16123 and 16329 were used, without dropping off insertions or deletions, as opposed to other studies (Richards *et al.*, 1996; 1998). However, more often than not, the obtained networks were very messy presenting high-dimensional cubes. As this is thought to be due to mutation rate heterogeneity in the control region, nucleotide

positions had to be weighted differently. In agreement with the findings of Wakeley (1993) and Hasegawa et al. (1993) and as used by Richards et al. (1998) and Helgason et al. (2000), polymorphic sites were divided into three groups according to their mutation rate: fast, average and slow (usually these latter ones correspond to a few coding-region RFLP data). To identify the fast mutating sites, an exploratory analysis was run. The statistics function of the programme showed the frequency of each segregating site. The positions with high frequency (implying high mutation rate) should be low-weighted in the input file for the subsequent analyses (Macaulay, pers.comm.). In this way, these sites are allowed to appear more than once in the in the network, and hence reducing the reticulations observed.

Sites of rapid change, namely 129, 172, 189, 192, 223 and 311 were given a low weight (*w*) of 5 and the remaining sites were assigned a weight of 10 in the input file for the programme. If the resulting networks were still difficult to interpret, the reduced median (RM) and median joining (MJ) were applied sequentially (Macaulay, pers.comm.).

To investigate whether a) previous results could be reproduced and b) resolution would be lost to the fewer segregating sites available, RMNs were first constructed based on the sequence data of Richards et al. (1996) both full length (270 bp) and trimmed to the sequence length amplified in this study (207 bp).

RMNs were constructed for the modern and ancient populations of England, and also for each ancient community separately. In addition, to visualise the continuity between ancient and modern mtDNA haplotypes, both populations were analysed together. Given the complexity of the network, RMNs were constructed for the major mtDNA clusters, namely JT, U5 and IXW, separately. For each of these major groups a separate input file was constructed including those modern and ancient HVS-I sequences sharing segregating sites 126, 270 and 223, which define clusters JT, U5 and IXW, respectively.

Chapter III - Optimisation of methods to retrieve poorly preserved DNA from hard tissues.

III.1 INTRODUCTION

The last decade of molecular research has proven that the analysis of DNA from ancient specimens is now possible. However, it requires a meticulous and thorough methodology which needs to be set-up prior to the relevant investigation. As archaeological specimens are often not abundant, preliminary studies can be undertaken in an equivalent material such as forensic samples.

III.1.1 DNA decay

When an organism dies, its DNA normally becomes degraded by endogenous nucleases. Under fortunate circumstances, such as rapid desiccation, low temperature or high salt concentrations, nucleases can themselves become destroyed or inactivated before all nucleic acids are reduced to mononucleotides. If this is the case, still slower processes start affecting the DNA. These involve hydrolityc and oxidative damage.

Nucleic acids undergo spontaneous decomposition in solution, RNA being particularly vulnerable, as the 2′-hydroxyl group of the ribose makes the phosphodiester bonds of RNA molecules very susceptible to hydrolysis. On the other hand, in the DNA molecule, the absence of the sugar 2′-OH group results in a more resistant phosphodiester bond but as a detriment, in a labile *N*-glycosyl group (Lindahl, 1993). Hydrolysis of the *N*-glycosyl bond and consequent release of a purine or pyrimidine base from the deoxyribose sugar of the DNA backbone, generates an abasic (apurinic or apyridimic) site (Wilson & Barsky, 2001). In addition, DNA bases are susceptible to hydrolytic deamination, which is the removal of the amino group form the nitrogenous base. By hydrolytic deamination cytosine is converted to uracil, adenine to hypoxanthine, guanine to xanthine and 5-methylcytosine to thymine, with cytosine and its homologue 5-methylcytosine being the main targets for this hydrolytic attack

(Lindahl, 1993). Both hydrolytic processes lead to destabilization and breaks in the DNA molecule.

DNA oxidation also represents another important source of damage, modifying the nitrous bases and the sugar-phosphate backbone of the DNA (Breimer, 1990; Lindahl, 1993). The major base lesion caused by hydroxyl radicals is 8-hydroxyguanine, which base-pairs with adenine (instead of cytosine) (Shibutani *et al.*, 1991).

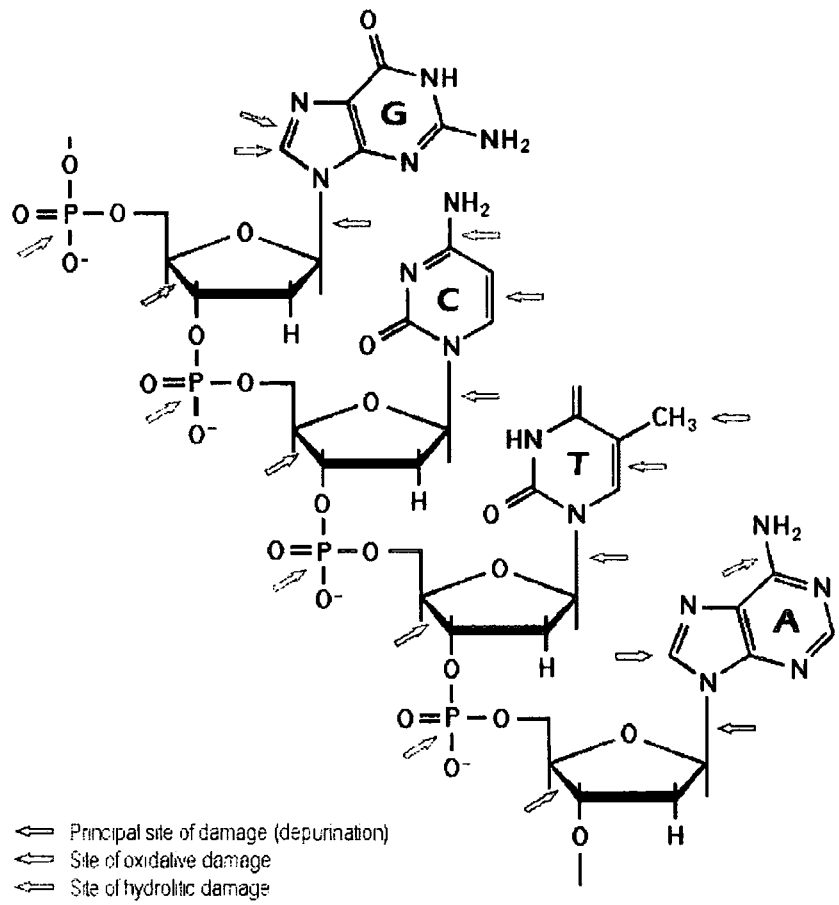


Figure III.1 Decay of DNA due to hydrolytic and oxidative damage. A short segment of one strand of DNA double helix is shown with the four bases. Principal sites of damage are indicated by black arrows. Sites susceptible to hydrolytic attack are indicated by grey arrows and those prone to oxidative damage by white arrows. (G): guanine; (C): cytosine; (T): thymine and (A): adenine (taken from Lindahl, 1993).

In-vivo, hydrolysis and oxidation are counteracted by specific DNA repair systems (Breimer & Lindahl, 1985; Tchou *et al.*, 1991; Dianov *et al.*, 1992). However, deprived of these repair mechanisms, DNA of non-living cells is heavily damaged. Ancient DNA molecules are thus broken into fragments of a couple of hundreds of base pairs (Pääbo,

1989; Handt *et al.*, 1994a; 1996; Francalacci, 1995; Colson *et al.*, 1997; Lindahl, 1997) and shows an abundance of non-repaired lesions (Hofreiter *et al.*, 2001).Very recently, the pattern of post-mortem damage has been fully characterised (Gilbert *et al.*, 2003a; 2003b). Although in principle this scenario may seem discouraging, finding characteristic ancient DNA behaviour (i.e. failure of amplification of long fragments and DNA artifacts) can be interpreted to support the authenticity of the aDNA (see Chapter IV).

III.1.2 Ancient DNA field

The extraction of ancient DNA (aDNA) from the mummified skin of a Quagga, an extinct member of the horse family (Higuchi, 1984), was the onset of the aDNA field. Since then, a large number of papers have reported the extraction of DNA from different ancient sources, such as soft and hard tissues (e.g. Hagelberg *et al.*, 1989; Merriwether *et al.*, 1994; Woodward *et al.*, 1994a), mummies (e.g. Pääbo, 1985; Handt *et al.*, 1994b), coprolites (e.g. Hofreiter *et al.*, 2000) and even papyri (e.g. Marota *et al.*, 2002) and paintings (e.g. Reese *et al.*, 1996). Unfortunately, during the early to mid-1990s a number of sensational and subsequently discredited reports describing the retrieval of DNA from fossilised plants (Golenberg, 1990; Cano *et al.*, 1993), ember-entombed insects (Desalle *et al.*, 1992) and dinosaur bones (Woodward *et al.*, 1994b) obscured the credibility of the field. As a consequence, researchers have concentrated their efforts in developing and perfecting appropriate methods for the isolation of genuine aDNA. By 1998, when I began this research, there was still debate regarding aDNA methods, and DNA extraction in particular. Therefore, it was thought that a preliminary methodological study would be necessary.

III.1.3 Ancient DNA and Forensic Sciences

Decay of DNA begins immediately after death. DNA from fixed or forensic samples also undergoes this diagenetic process, and behaves in a similar fashion to archaeological material as a result of the severe fragmentation of their molecules (Capelli *et al.*, 2003). Both molecular archaeology and forensic genetics use PCR technology to overcome the problem of expanding the surviving molecules (Pääbo,

1989; Handt *et al.*, 1994a; Capelli *et al.*, 2003). They also share a number of experimental difficulties, such as the presence of inhibitory molecules, the risk of contamination and the optimisation of amplification conditions. Whereas analyses of DNA from archaeological specimens allow investigation of the genetic structure of past populations and its implications in evolutionary process, analyses of DNA from forensic cases can help reveal more recent historical events.

Genetic analyses have allowed the identification of several 'famous' individuals, such as Louis XVII (Jehaes *et al.*, 2001), the Romanov and the Prince Branciforte Barresi families (Gill *et al.*, 1994; Rickards *et al.*, 2001), Martin Bormann (Anslinger *et al.*, 2001) and Josef Mengele (Jeffreys *et al.*, 1992); and also the identification of anonymous people, such as war fatalities (Holland *et al.*, 1993) and massacre victims (Primorac *et al.*, 1996; Vollen, 2001), of which many more remain unidentified.

III.1.3.1 Identification of skeletal remains from mass graves

During the 1970s, several South American countries were shaken by periods of intense violence, as ruling dictatorships committed serious human rights abuses. That was the case in Argentina, where in 1976 a military coup seized power. Their alleged aim was to defeat the increasing far-left extremist wave. However, not only activists but anyone suspected of political opposition of any kind were persecuted. Suspects would be abducted by military forces and taken to concentration camps where they would be interrogated in torture sessions and eventually murdered. The resulting corpses were disposed of in anonymous mass graves or dumped in the sea from planes. As neither the bodies nor any information on their whereabouts or fate was ever given to their families, they are referred to as the 'disappeared'. In Argentina, it is thought that this has been the dreadful destiny of at least 10,000 people (CONADEP, 1984).

Many cemeteries have been used by Argentinean military forces for illegal purposes. However, very little official information about the location of burial sites is available. When they do exist, cemetery files are generally incomplete or false. The Equipo Argentino de Antropología Forense (EAAF), the human rights organization in charge of exhumations and historical investigation, has discovered and unearthed several mass graves. The largest contained as many as 380 corpses. Most bodies in these graves show

cranial fractures caused by being shot at close range. This supports the hypothesis that they are the unidentified remains of murdered people, very likely *disappeared* political prisoners (Cohen Salama & EAAF, 1992).

To identify these remains, the EAAF has traditionally used an approach based on comparisons of osteological and anatomical analyses which consists of two stages. On one hand, the exhumed skeletons are examined to determine sex, stature and age at time of death (e.g. Albert, 1998; Mall *et al.*, 2001). Cause of death can be established by ballistic analysis. Bone injuries and dental features are also recorded. On the other hand, the physical characteristics of *disappeared* persons are collected through interviews with their relatives. Surgeries that might have left unequivocal bone scars, X-rays and pictures where dental features are clearly displayed are of special interest (e.g. Brkic *et al.*, 2000; Pretty & Sweet, 2001; Kirk *et al.*, 2002). Based on relevant historical information, hypotheses about the identities of the victims are established. Thus, the skeletal remains are matched up with their putative relatives and pre-mortem and post-mortem physical characteristics are compared. If the data match, the victim is positively identified. However, not many bodies could be identified by this method, as the compiled information was often not conclusive enough. Therefore a different strategy to identify a greater proportion of the remains was needed. In fact, in 1997, the first genetic identification of one of these cases was carried out (Corach *et al.*, 1997).

III.1.3 Aims

This chapter will describe the development of the methods required for the work on poorly preserved DNA from skeletal material. This extends from the initial decomposition of the hard tissue to the final DNA sequencing. As archaeological and forensic materials are believed to behave in a similar manner, the methodology will be set-up first in forensic samples. In addition, the results will be used to solve a forensic case: the identification of human skeletal remains victims of human rights abuses.

III.2 Results and Discussion

III.2.1 Forensic material and preparation of samples

Skeletal remains from individuals exhumed from a number of mass graves excavated in Buenos Aires, Argentina were used for this preliminary study (for details see Appendix B). The forensic material was 22 year-old, of which 10 years had been buried in soil (without coffin) and 12 years stored in open air. Preservation state varied considerably among samples.

Teeth were chosen as a tissue for the DNA extraction due to their unique histological and anatomical organization. Dental hard tissue physically encloses the pulp, offering a structure of great durability. Dentine and enamel provide a protective enclosure, which prevents degradation due to microbial attack (Smith *et al.*, 1993). Contamination of inner tissues in-situ or during excavation and curation is unlikely given the impermeability of the enamel. Moreover, teeth are the best-preserved skeletal element in the archaeological and forensic record (DeGusta & White, 1996).

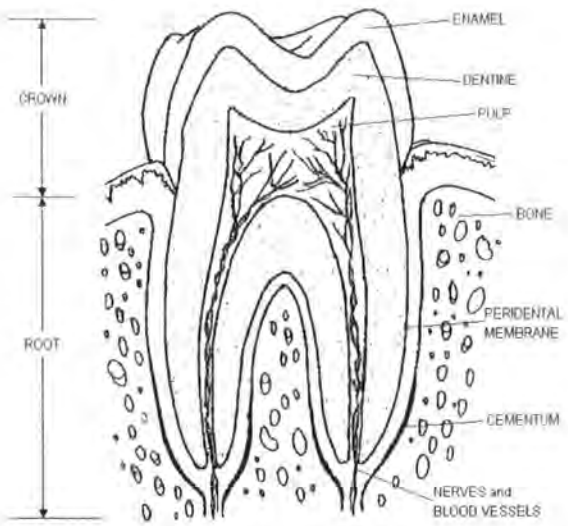


Figure III.2 Anatomic structure of a tooth. DNA is likely to be found in the pulp or adjacent areas.

The soft tissue within the pulp cavity consists of odontoblasts, fibroblasts, endothelial cells, periphereal nerve cells and nucleated components of blood (Frick *et al.*, 1991). Hence, the coronal pulp chamber is the main target for DNA sampling in

forensic case-work. Cell remnants of the pulp might remain attached to the chamber walls, allowing recovery by grinding (Drancourt *et al.*, 1998), but generally pulp tissue is absent in most archaeological samples. Hard dental tissue should therefore be considered as an additional source of DNA. Dentine matrix contains the same chemical constituents as osseous tissue –which was the first hard tissue used for aDNA extraction (Hagelberg *et al.*, 1989), but in different ratios. Hydroxiapatite, the compound that binds and thus preserves DNA is present at a higher concentration than in bone, making dentine potentially a better source of ancient DNA (DeGusta *et al.*, 1994). In addition, DNA has successfully been extracted from teeth subjected to various environmental conditions and exposure periods (Schwartz *et al.*, 1991; Pfeiffer *et al.*, 1999) whereas bone seems to be more susceptible to them (Waite *et al.*, 1997). Thus, teeth were preferred to skeletal material and both pulp cell remnants and dentine were used as source tissue for the DNA extraction.

Two protocols were tested in order to obtain fine-powdered dental material suitable for the DNA extraction (in both cases teeth were previously cleaned as described in Methods II.2.5):

The "drill in" approach – as has been already described in Methods II.2.5, and

The 'nut & bolts' approach (Thomas & Moore, 1996) – this low-tech system consisted of two stainless steel bolts assembled with one stainless steel nut (all previously cleaned with bleach and autoclaved). One of the bolts was removed and the dental or bone sample placed inside the nut. The bolt was placed back and screwed until the sample broke (see Fig. III.3). To remove the dental pieces, one of the bolts was unscrewed and the other bolt screwed further into the nut, pushing the sample out. To avoid cross-contamination between samples, nuts and bolts (n&b) were used only once.

Figure III.3 Scheme of the *nut & bolt* apparatus. The tooth is placed inside the nut and then the bolt screwed in so that the dental sample breaks.

Two 'n&b' apparatus sizes were used (of 5 and 30 mm ∅). First, the entire tooth was placed inside the bigger n&b and broken up in several pieces. The small fragments were then transferred into the smaller apparatus so that the pieces of tooth could be ground into a fine powder. The fine powder was collected into an eppendorf tube. This method had the advantage that minimum material was lost or disseminated during the procedure. However, the powdered samples consisted of a mixture of dental material (i.e. pulp remnants and dentine). Pulp as such is rarely present in forensic material yet its remnant cells have been reported to present the largest amounts of DNA (Drancourt *et al.*, 1998). Therefore, the drill was preferred as allowed the collection exclusively of the most inner material, attached to the pulp chamber walls. Separate fractions of dentine were also collected. For the forensic samples, DNA could be isolated from either fraction. However, for the archaeological samples, more often than not DNA isolation from the first fraction of material collected (i.e closer to the pulp) yielded better results than dentine alone (see Chapter IV and Appendix C).

III.2.2 Optimisation of aDNA extraction methods

A large number of protocols for the isolation of ancient DNA have been described, such as the standard phenol:chloroform (Hagelberg & Clegg, 1991), silica binding (Boom *et al.*, 1990; Hoss & Pääbo, 1993) silica-based columns (Greenspoon *et al.*, 1998; Yang *et al.*, 1998), salting out precipitation (Cattaneo *et al.*, 1995), resin separation (Sweet *et al.*, 1996), enzymatic treatment (Pusch & Scholtz, 1997), streptavidin-coated magnetic beads hooking (Tofanelli *et al.*, 1999) and even voltage-induced (Bachmann *et al.*, 2000) methods. They all have been widely applied for DNA extraction from forensic samples and archaeological material. Most of them reported good, yet highly variable results.

Seven DNA extraction methods, some of them commercially available kits, were investigated in this preliminary analysis. Replicate samples of approximately 80 mg of dental powder, obtained by drilling the tooth as described in Methods, section II.2.5, were subjected to different DNA extraction protocols, namely GeneClean (Bio Rad), QIAamp tissue kit (QIAgen), QIAquick PCR purification kit (QIAgen), Silica I, Silica II, InstaGene™ (Bio Rad) and standard phenol:chloroform (for details see Appendix A).

For this, 8 teeth were used (samples D10, D11, D12, D13, D14, D15, M6 and D8-42), since a maximum of four replicates could be obtained from each dental sample. The yield of DNA was not quantified, as very small amounts of degraded genetic material were expected, which have been reported to give anomalous A260/A280 measurements (Previdere *et al.*, 2002). Instead, the resulting DNA solutions were amplified by PCR as a mean to assess the efficiency of the extraction method. Quantification of amplification products was measured in relative terms, based on the degree of brightness of the bands in the gel. Not only the amount of aDNA retrieved, but also other factors, such as the number of manipulation steps, duration of protocol and toxicity of the reagents were taken into consideration when choosing the extraction method. Figure III.4 shows the results of one trial for the seven extraction protocols.

InstaGene™ is a simple and straightforward commercially available method (Bio Rad Ltd.). It consists of a 5% solution of a chelating resin (Chelex®) which works by lysing the cell membrane and releasing the DNA out into solution. The matrix also adsorbs cell lysis by-products that may interfere with the PCR amplification process, thus purifying the DNA. Chelex® has been widely used for DNA isolation from forensic specimens as saliva, blood and semen stains (Walsh *et al.*, 1991; Woodward *et al.*, 1994a; Sweet *et al.*, 1996). However, the method was not sensitive enough for extracting DNA from forensic hard tissue samples, as can be seen in Figure III.4 (lower gel), set (f) of samples. This was the only protocol which did not allow the amplification of DNA from any forensic sample. The efficiency of the extraction could not be improved with longer incubation at room temperature or 56 °C (not shown).

The original silica-based method was described by Boom et al. (1990). The method is based on the lysing and nuclease-inactivating properties of the guanidinium thiocyanate together with the nucleic acid binding property of silica particles. In presence of high concentrations of caotropic agents such as NaI, $NaClO_4$ or GuSCN, DNA will bind to diatoms, silica or glass particles, forming a complex that can be pelleted by centrifugation. Based on this method, several other protocols were developed specially for aDNA work, some of them commercially distributed. The GeneClean for ancient DNA is a commercially available kit (Bio 101 Inc.) designed for isolation of DNA from bone, preserved tissue, soil organisms, etc. Even though the exact composition of

solutions is not available for public domain, it is known this is a silica-based method, in which the silica is loose rather fixed to the filter membrane. The two QIA methods tested are modifications of commercial kits (QIAgen) based on the same principle, although in this case the silica is fixed to the filter membrane in the spin filter columns. The former was designed for purification of DNA from most commonly used human tissue samples, such as muscle, liver, heart and brain. The latter is used for purification of PCR products.

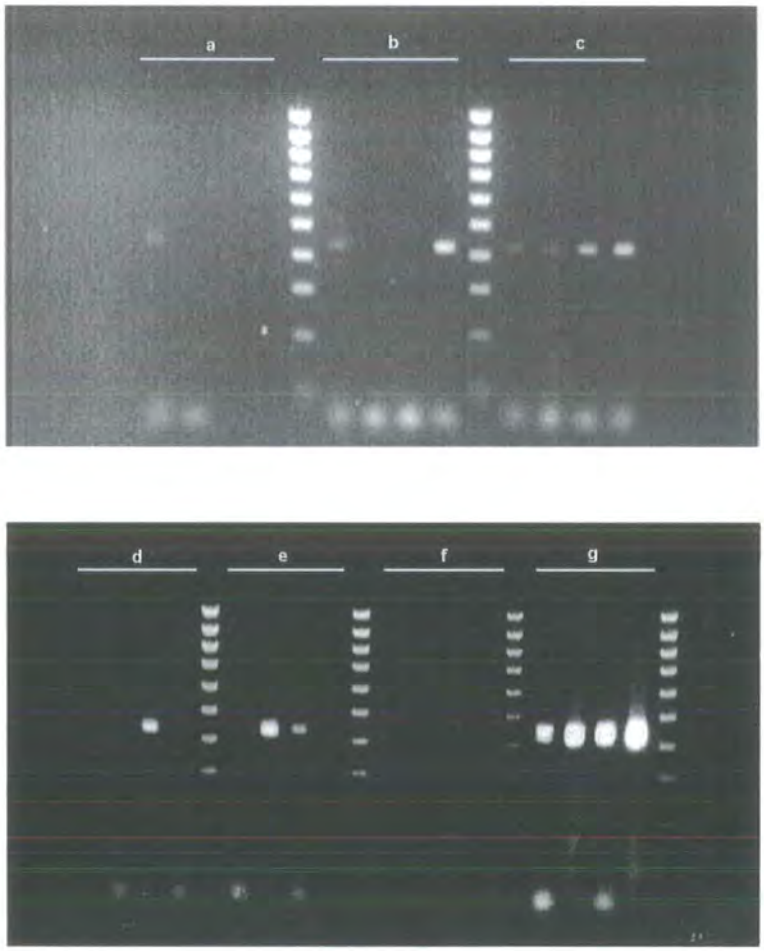


Figure III.4 Assessment of DNA extraction protocols. Upper gel: DNA extraction success for three protocols: a) GeneClean for ancient DNA (Bio101 Inc.), b) QIAamp tissue extraction kit (QIAgen), and c) QIAquick PCR purification kit (QIAgen) tested in four samples (D10, D11, D12 and D13, lines 1 to 4 of each method). Lower gel: DNA extraction success for four protocols: d) Silica I, e) Silica II, f) InstaGene matrix (Bio Rad), and g) standard phenol extraction tested in four forensic samples (D14, D15, M6 and D8-42, lines 1 to 4 of each method). A 100-bp ladder was run between the extraction methods.

Both QIAgen methods allowed isolation of DNA from the forensic samples, as can be seen in Figure III.4 (upper gel) sets of samples (b) and (c), with only the QIAquick protocol working for all the samples. On the other hand, the GeneClean for aDNA kit yielded only a very weak PCR product (Fig. III.4 (upper gel), set (a) of samples). The manufacturer warns that a severe reduction of recovery for DNA fragments between 100 and 500 bp may occur and some small fragments may not be recovered at all (Bio 101, 2001). To overcome this problem, the DNA-Glassmilk binding step was performed at 60 °C as suggested by Smith *et al.* (1995).

It has to be noticed that only broad conclusions may be drawn from the comparison of protocols. Whereas the Instagene™ method seemed to be not appropriate for extracting DNA from poorly preserved hard tissue samples, efficiency of the other methods should be analysed with caution. As can be seen in Figure III.4, the same extraction protocol gave different results for the different dental samples, which may be due to the amount of DNA present in the collected dental powder. Likewise, further repeat trials also yielded heterogeneous results (i.e. the same method performing differently with different samples). DNA extraction should be read in qualitative terms, as to whether or not an extraction method allowed the isolation of DNA from these poorly preserved hard tissues.

The standard phenol:chloroform protocol yielded the best results, both in terms of the number of positive reactions and the relative quantity of DNA amplified as indicated by the brightness of the bands (see Fig.III.4 (lower gel) set (g) of samples). However, since the reagents involved are toxic, a less harmful protocol was preferred instead. The QIAquick PCR purification kit method described by Yang *et al.* (1998) seemed to be a more suitable protocol, as proved to be appropriate for aDNA work and avoided toxic reagents. Thus, the QIAquick purification kit was used as the standard protocol for the isolation of DNA from the forensic samples. In this way it has been possible to isolate DNA from all the 32 twenty-two-year-old skeletal remains.

However, when three test archaeological samples from the Norton site were subjected to this extraction protocol, no aDNA was isolated. Only one PCR product was obtained which turned out to be a contaminating sequence. Inhibition of PCR was not detected and no improvement was obtained by increasing the number of amplification

cycles or the amount of DNA solution in the reaction (see below). Hence more sensitive methods were necessary. As the number of test archaeological samples was very limited, replicates could not be done at this stage. A combination of the organic extraction (which also yielded the best results for the forensic samples) and a silica-based concentration was chosen (modified from Richards & Sykes, 1995). For the cell lysis several buffers were tested containing different EDTA concentrations and dithithreitol or spermidine, which have been reported to improve the extraction of DNA from for osseous and dental samples (Merriwether *et al.*, 1994; Montiel-Duarte, 2000; Harvey, pers.comm.). Improvement of the DNA-silica binding step was achieved by modifications based on the temperature (Smith *et al.*, 1995) and pH (Schmerer *et al.*, 1999) dependence of this process. A commercial kit was used as it allowed checking the pH of the solution. Description of the final protocol used for isolation of DNA from archaeological material was given in Chapter II.2.6).

III.2.3 Optimisation of PCR amplification

III.2.3.1 Primer design and PCR product length

Mitochondrial DNA (mtDNA) has been widely used in forensic casework due to its high variability and large number of copies per cell. Primers for amplification of the control region previously published by Vigilant et al. (1989) flank a 540 bp and a 420 bp fragments of the HVS-I and HVS-II, respectively. However, fragments of such length are unlikely to be amplified from degraded DNA templates. Thus, a number of European, Asian and African mtDNA sequences were aligned to identify other possible priming regions. New primers were designed using Oligo 4.0 software (National Biosciences Inc.) so that their different combinations would target fragments of different lengths. Primers sequences are listed in Table III.1 and their relative location in the control region is shown in Figure III.5. Lengths of the resulting PCR products are indicated in Table III.2. Primer and Mg^{2+} concentration as well as annealing temperature was titrated using a positive aDNA extract.

Primer	Sequence	Reference
HVS-I		
L15997	5'-CACCATTAGCACCCAAAGCT-3'	Vigilant *et al.*, 1989
L16099	5'-AACCGCTATGTATTTCGTAC-3'	this work
L16179	5'-CATAAAAACCCAATCCACATC-3'	this work
L16209	5'-CCCCATGCTTACAAGCAAGT-3'	Montiel, 2000
L16348	5'-GTACATAGCACATTACAGTC-3'	this work
L16517	5'-CATCTGGTTCCTACTTCAGG-3'	this work
H16331	5'-TTTGACTGTAATGTGCTATGTA-3'	this work
H16498	5'-CCTGAAGTAGGAACCAGATG-3'	Vigilant *et al.*, 1989
HVS-II		
L29	5'-GGTCTATCACCCTATTAACCA-3'	Vigilant *et al.*, 1989
L97	5'-CGCGATAGCATTGCGAGACG-3'	this work
H78	5'-CGTCTCGCAATGCTATCGCG-3'	this work
H251	5'-GGAAAGCGGCTGTGCAGAC-3'	this work
H313	5'-GTGCTGTGGCCAGAAGCGGG-3'	this work
H408	5'-CTGTTAAAAGTGCATACCGCA-3'	Vigilant *et al.*, 1989

Table III.1 Sequences of the oligonucleotides used for PCR amplification. Numbering is according to Anderson *et al.*, 1981. (H): heavy strand and (L): light strand of DNA (i.e. forward and reverse primers, respectively).

Combination of primers			PCR fragment
HVS-I			
L15997	→	H16498	541 bp
	→	H16331	376 bp
L16099	→	H16498	439 bp
	→	H16331	264 bp
L16179	→	H16498	360 bp
	→	H16331	195 bp
L16348	→	H16498	190 bp
HVS-II			
L29	→	H408	422 bp
	→	H313	324 bp
	→	H251	263 bp
L97	→	H408	352 bp
	→	H313	256 bp
	→	H251	193 bp
Inter-region			
L16517	→	H78	170 bp

Table III.2 Combination of primers used to PCR amplify several fragments of the hypervariable control regions I and II. Length of the obtained PCR fragments (including primers) is indicated in base pairs (bp).

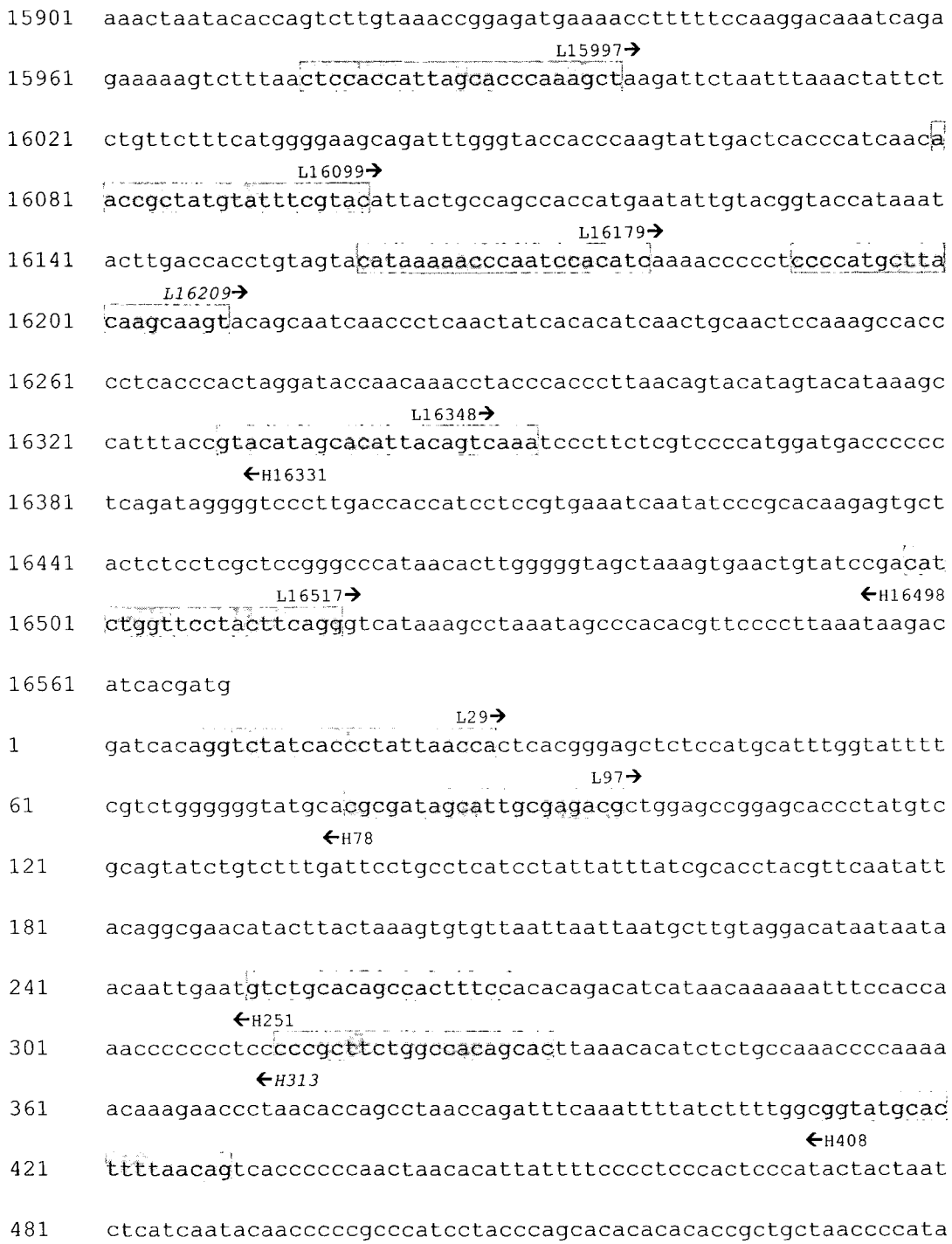


Figure III.5 Location of the HVS-I and HVS-II primers tested for PCR amplification. Numbering is according to Anderson et al., 1981. Sections [15901-15569] and [1-540] correspond to the HVS-I and HVS-II, respectively. For overlapping oligos, forward primers (L) are indicated on the top side of the sequence and reverse primers (H) on the bottom side. Arrows indicate direction of amplification. Sequencing primers are indicated in *italic*.

Different combinations of primers were tested. As expected, long PCR fragments (> 400 bp) were difficult to amplify from the forensic material and not amplifiable at all from the archaeological samples (see Fig. III.6). On the other hand, amplification of short PCR fragments (<200 bp) proved to be highly prone to contamination of PCR and extraction blanks (see Fig. III.7) which is likely due to contaminating DNA in plasticware disposables.

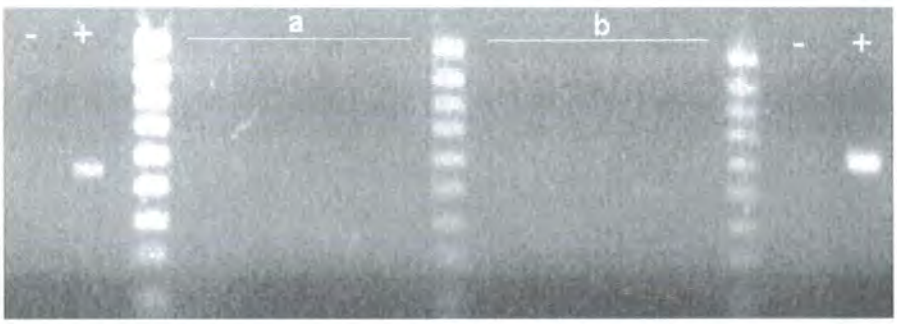


Figure III.6 Test for PCR amplification of HVS-I and HVS-II fragments using primers originally described by Vigilant *et al.*, 1985: **a**) L15997/H16498, and **b**) L29/H408.

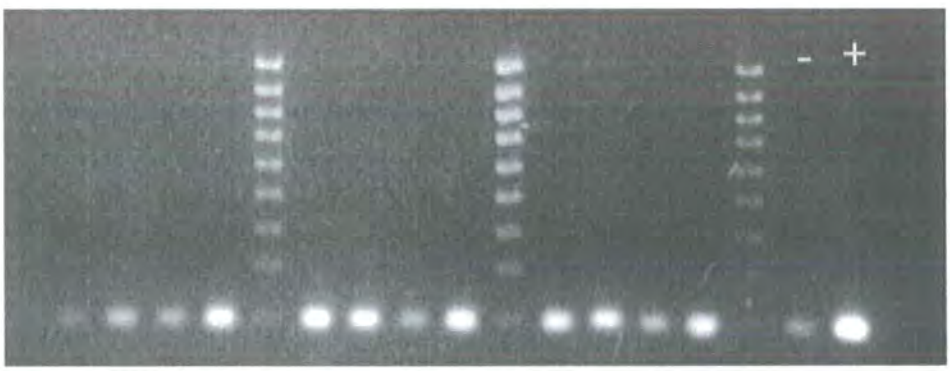


Figure III.7 DNA contamination of PCR and extraction blanks detected during PCR amplification of a short fragment using primers L16348 and H16498.

As a test, amplification of a short fragment was carried out in two batches of PCR tubes: with and without UV-treatment. Figure III.8 shows that more and brighter PCR-product bands were obtained for the samples set-up in non-UV tubes, suggesting the presence of contaminating DNA. Contamination in plasticware has been reported to be (inversely) correlated with the size of the PCR fragment amplified (Schmidt *et al.*, 1995). UV-irradiation of disposable plasticware was carried out as a normal routine.

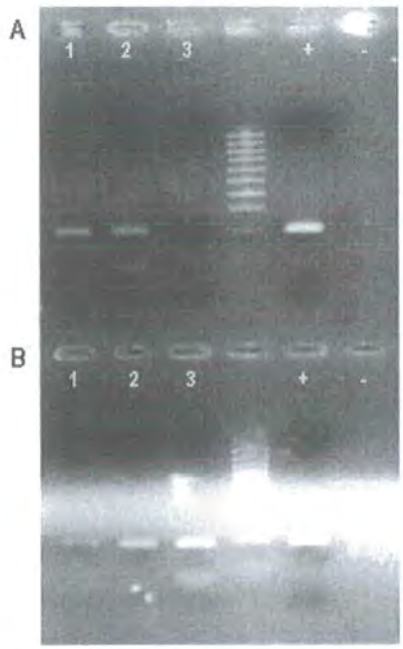


Figure III.8 Test for contamination in plasticware and UV-treatment as proposed by Sarkar and Sommer (1993). A short sequence (primers L16348 H16498) was amplified from three forensic samples (1 to 3). PCR reactions were set-up in tubes **A)** with, and **B)** without UV-treatment.

Fragments of intermediate size, between 200 and 300 bp proved to be amplifiable from the forensic and also from the archaeological samples (see Fig. III.9) and seemed to be less prone to contamination with spurious short DNA fragments. Although contamination was still an issue and had to be controlled and monitored, PCR and extraction controls were usually clean. Thus, primers [L16099/H16331] and [L97/H313] were used to amplify HVS-I and HVS-II, respectively. In addition, these fragments comprise the most variable sectors within HVS-I and HVS-II (Baasner *et al.*, 1998; Lutz *et al.*, 1999; Meyer *et al.*, 1999) so that exclusion power and resolution were not diminished.

Figure III.9 PCR amplification of medium size (270 bp) fragments using primers L16099/L16331. Sample (*) shows a possible case of inhibition as suggested by the lack of primer-dimer (see text).

III.2.3.2 Overcoming inhibition of the amplification reaction

Organic and inorganic compounds that inhibit the amplification of nucleic acids are common contaminants of DNA samples from various sources. They can interfere with the PCR reaction at several levels, leading to different degrees of attenuation and even to complete inhibition. A wide variety of PCR inhibitors have been reported and they appear to be especially abundant in complex samples such as animal fluids, food and organic soils (Wilson, 1997). They are also frequently found in extracts from ancient samples (Horai, 1989; Pääbo *et al.*, 1989a; Hagelberg & Clegg, 1991; O'Rourke *et al.*, 1996). The exact nature of these particular inhibitory compounds is still unknown. Pääbo (1989) proposed that they are related to Maillard products generated during the reduction of sugars of decomposing tissues. This seems to be particularly relevant to soft tissue extracts. Regarding DNA extracts from skeletal remains, the inhibitory compounds might be soil components such as iron, tannins or humic acids (Hagelberg & Clegg, 1991), which are also found in environmental samples (Abbaszadegan *et al.*, 1993). They exhibit a similar solubility to DNA and as a consequence are not completely removed during standard purification procedures (e.g. phenol/chloroform) (Moreira, 1998).

DNA can be separated from these inhibitory compounds by electrophoresis in low-melting-point agarose gel (Moreira, 1998), by using gel filtration resins (as Sephadex type) (Miller, 2001) by nitrocellulose membrane filtration (Kiesslich *et al.*, 2002) or by caesium chloride density gradient centrifugation (Juniper *et al.*, 2001). However, none of these methods is suitable for ancient DNA extracts, as the small number of templates present in the solution might be completely lost during any these procedures. Instead a strategy that involved fewer steps and was therefore less wasteful would be necessary.

Although lack of PCR amplification can simply be due to absence of DNA, PCR inhibition can be clearly identified. If adding an aliquot of aDNA extract to a positive DNA control prevents amplification, then inhibition of the DNA polymerase is confirmed. This was the case for some of the negative amplification reactions obtained. Figure III.10 shows an example. Sample D5 was thought to present inhibitory compounds. In fact, when this sample was added to a positive DNA, no amplification was obtained. Failure of amplification was not due to an excess of forensic aDNA (see

Fig. III.10, lane 3). It has been suggested that inhibition of PCR could be caused by large amounts of degraded material (Kolman & Tuross, 2000).

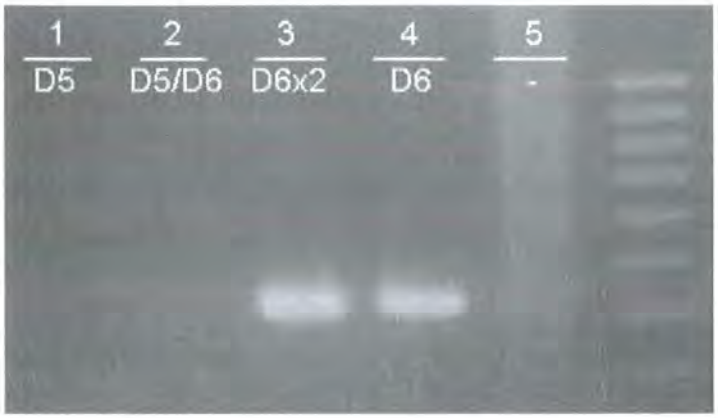


Figure III.10 Lack of amplification reaction de to inhibitory compounds. Lanes 1 to 4 show the amplification results of 25 µl-PCR mixes for: (1) 5 µl of forensic sample D5, thought to present inhibitory compounds; (2) 5 µl same forensic sample *and* 5 µl positive forensic sample (D6); (3) 10 µl forensic sample D6; (4) 5 µl of forensic sample D6 and (5) negative control.

Inhibition was found to be associated with absence of primer-dimer formation (see previous Fig. III.9, sample indicated with *). This can be explained by the mechanism for primer-dimer formation. This process requires the two primers, the dNTPs and an *active* DNA polymerase enzyme as the starting reactives (Das *et al.*, 1999). The presence of strong inhibitory compounds will hinder all activity of the polymerase enzyme, resulting in absence not only of amplification products but also of these by-products of the reaction. This fact was indirectly used to identify inhibition in the samples.

To overcome the inhibition of amplification several strategies were undertaken:

Dilution of samples – DNA extracts that yielded a negative PCR result were diluted ½ to 1/20 in sterile water and re-amplified under standard conditions.

Blotto method – samples yielding negative PCR results were re-amplified adding 2% (v/v) BLOTTO (10% skim milk powder and 0.2% NaN_3) to the reaction (DeBoer *et al.*, 1995). For some samples a range of 1-5% (v/v) BLOTTO was used.

Purification with Chelex – samples giving a negative amplification result were further purified by Chelex resin. 50 µl of 5% Chelex were added to the DNA solutions. Samples were incubated for 30 min. at 65 °C and 10 min. at 95°C. The resin was

pelleted by centrifugation at high-speed so that the DNA was ready for a re-amplification. As the solutions were diluted ½ during this procedure, 10 µl instead of 5 µl were used for the PCR reaction

Addittion of BSA – titration of the optimum bovine serum albumin (BSA) concentration was carried out for an inhibitory extract, using 0.5 to 4 µl of BSA 2 µg/µl in 25-µl reactions.

Cold-step – the method simply consists in storing the DNA extraction solutions at 4 °C at least for three days and then at –4 °C until the PCR reaction is performed.

The simplest approach to overcome inhibition of the amplification reaction is to dilute the DNA extract to be used in the reaction. This decreases the amount of inhibitory compounds, allowing amplification. However, it may also reduce the amount of aDNA below the level of detection (Francalacci, 1995). In 11 of 18 forensic samples that inhibited the PCR, this could be overcome by diluting the extraction prior to the enzymatic reaction. Optimal dilutions ranged from ½ to 1/20 and varied according to the sample. However, if extracts were further diluted, less or no amplification was observed (e.g. see Fig. III.11).

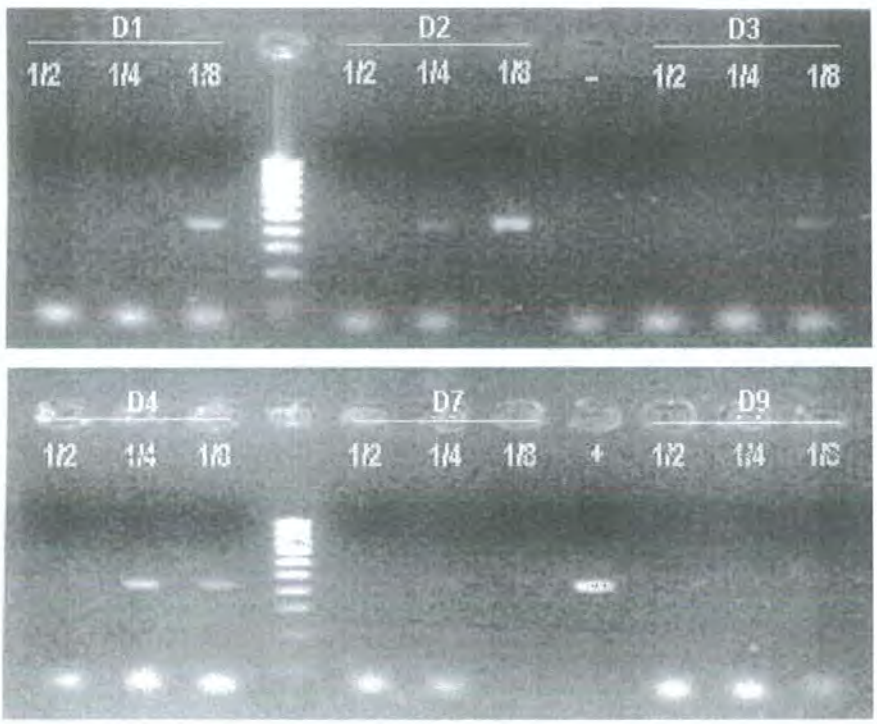


Figure III.11 Overcoming PCR inhibition by dilution of the DNA extracts. Results obtained for 6 samples (D1, D2, D3, D4, D7 and D9) diluted ½ to 1/8 in sterile water.

Non-fat milk solutions, termed BLOTTO (Bovine Lacto transfer technique Optimizer) (Johnson *et al.*, 1984), have been widely used to prevent non-specific binding of proteins and nucleic acids to nitrocellulose in Western and Southern blotting procedure. It was also reported that non-fat milk eliminated the inhibitory effect in plant extracts (DeBoer *et al.*, 1995). Furthermore, adding BLOTTO to extracts of ancient Sorghum DNA reduced the PCR inhibition (Deakin, pers. comm.). However, when 1 to 5% (v/v) BLOTTO was used with DNA extraction that showed PCR inhibition, no improvement was observed. On the contrary, yield of positive PCR products was reduced with BLOTTO (see Fig. III.12).

Figure III.12 Test for alleviation of inhibition by BLOTTO. Suffix (B) indicates BLOTTO treatment of both the forensic samples (16, 17 and 18) and the negative and positive controls. These latter were carried out to detect contamination in BLOTTO reagents and effect of BLOTTO on positive samples.

BSA is a protein known for its capacity for stabilising enzymes. It increases the efficiency of the PCR reaction by facilitating the action of the DNA *Taq* polymerase during PCR. Furthermore, BSA might relieve inhibition of amplification by binding to inhibitory compounds present in the reaction, although the exact mechanism remains unknown (Kreader, 1996). The optimum quantity of BSA varied according to the sample, but large amounts of the protein were found to hinder the reaction (see Fig. III.13). Overall, a BSA 0.08 μg/μl final concentration seemed to give the best results. This amount of BSA was used as standard reagent of the aDNA PCR reactions, regardless of whether or not the extract presented inhibitory compounds.

Figure III.13 BSA titration for a forensic sample. Lanes 1-5 are decreasing concentrations of BSA: 0.32, 0.24, 0.16, 0.12 and 0.08 µg/µl, respectively.

InstaGene™ has originally been designed for DNA extraction procedures. The chelating resin also adsorbs the cell lysis by-products as well as other substances. It has been reported that InstaGene™ can be used to relieve inhibition of the amplification reaction (Sweet *et al.*, 1996; Montiel *et al.*, 1997; MacHugh *et al.*, 2000). Although this method did not allow isolation of aDNA, it did prove to be useful for the removal of inhibitory compounds. Inhibition was overcome in 5 of 8 samples that gave negative PCR results. However, any trace of the resin itself was found to be a strong PCR inhibitor (MacHugh *et al.*, 2000) (see Fig.III.14). To prevent this, samples were centrifuged at high speed to pellet the resin and the supernatant transferred to a clean new tube.

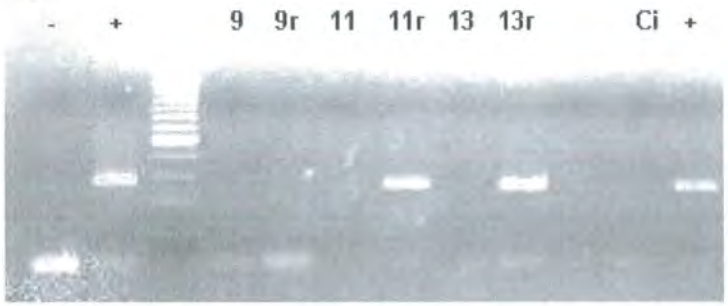


Figure III.14 Overcoming PCR inhibition by InstaGene™ resin. Suffix (**r**) indicates resin treatment of the forensic sample. A control for inhibition was also tested (Ci): a positive sample (following lane +) was treated with InstaGene and traces of the resin were intentionally left in the final solution.

The so-called 'cold-step' (Montiel *et al.*, 1997) was also attempted. The approach simply consists in storing the extracts at 4 °C at least for three days *before* the amplification reaction. For the PCR, an aliquot of this DNA solution is taken without

shaking or stirring the extract. It is also recommended not to touch the walls or the bottom of the tubes (Montiel *et al.*, 1997). It was observed that this procedure improved the PCR efficiency for some samples (see Fig. III.15). It had also been noticed previously that PCR reactions performed straight after the DNA extraction procedure usually yielded negative results. Therefore, a *cold-step* was always performed. Although the process involved is still unclear, it has been proposed that the cold temperature promotes formation of micelles or other phenomena such as crystallization or binding of the inhibitors to the wall of the polypropylene tube (Montiel *et al.*, 1997; Montiel-Duarte, 2000).

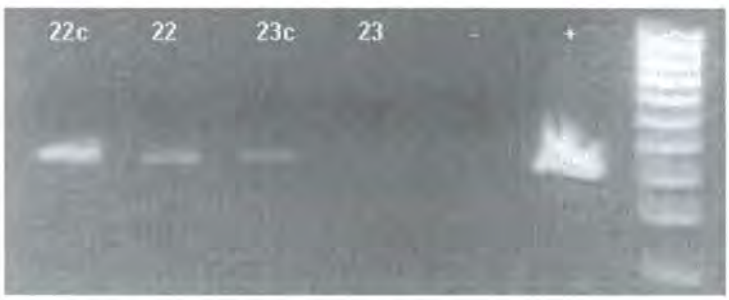


Figure III.15 Improving PCR amplification by the cold-step method. Suffix (**c**) indicates cold-step treatment of two forensic samples. Sample 22 shows improvement of amplification yield and sample 23 shows alleviation of the PCR inhibition.

III.2.3.3 DNA polymerase and reaction buffers

Ampli*Taq* Gold™ DNA polymerase was chosen for the amplification as it has the advantages of having a low error rate and being inactive at room temperature. The enzyme needs of an 11-minute activation step at 94 °C to become active, making the system a 'Hot Start' (Erlich *et al.*, 1991) PCR. This is particularly important to avoid primer-dimer formation during the PCR set-up and to increase the specificity and yield of the reaction (Moretti *et al.*, 1998). Although not thoroughly tested, it was noticed that the GeneAmp® 10x PCR Gold Buffer yielded better results than the standard one, the difference between them being the pH of the solutions (8.0 and 9.0, respectively).

III.2.3.4 Number of amplification cycles

The strategy reported by Rameckers et al. (1997) to determine the necessary number of amplification cycles, i.e. the number of cycles leading to detectable amounts of PCR products, was followed. The equation used to describe the PCR process is $y = a(1+F)^n$, where a = concentration of DNA targets, F = efficiency of PCR amplification (0.0-1.0), n = number of amplification cycles and y = detection limit. The number of cycles required can be calculated as $n = \frac{\ln(y/a)}{\ln(1+F)}$.

Based on this formula and given y = 10 ng, the number of necessary cycles were calculated (see Table III.3). This amount of PCR product was chosen in order to recognize signals from a 5 µl aliquot of 25 µl reactions in a standard ethidium bromide stained agarose gel and assuming a 200 bp PCR fragment. Given that for aDNA the PCR efficiency is estimated to be <0.5 and the number of DNA templates <1,000 (Rameckers *et al.*, 1997), 45 cycles were performed.

F	Targets					
	1	10	100	10^3	10^4	10^5
1.00	34	30	27	24	20	17
0.95	35	32	28	25	21	18
0.90	36	33	29	26	22	18
0.85	38	34	30	27	23	19
0.80	40	36	32	28	24	20
0.75	42	38	33	29	25	21
0.70	44	40	35	31	27	22
0.65	46	42	37	33	28	23
0.60	49	45	40	35	30	25
0.55	53	48	43	37	32	27
0.50	57	52	46	40	35	29
0.45	62	56	50	44	38	31
0.40	69	62	55	48	42	35
0.35	77	70	62	54	47	39
0.30	88	79	71	62	53	44
0.25	104	93	83	73	62	52
0.20	127	114	102	89	76	64
0.15	165	149	132	116	99	83
0.10	242	218	194	170	145	121

Table III.3 Number of necessary cycles to obtain detectable amplification product when starting from a small template number. (F) efficiency of PCR reaction (taken from Rameckers *et al.*, 1997)

A test for the number of PCR cycles and its relation with target PCR size is shown in Figure III.16. Short fragments of 90 and 175 bp could be obtained by 35-cycle amplifications whereas medium size targets needed of 45 cycles to be amplified. It remains unknown whether the amplification of short fragments is real or contaminating DNA.

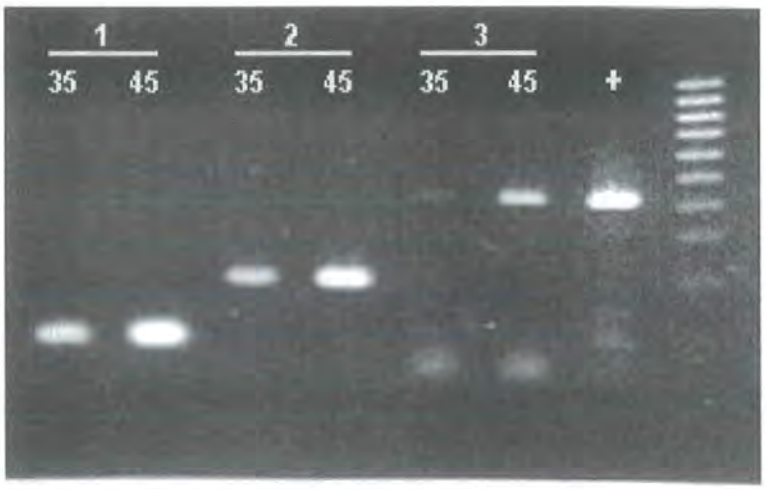


Figure III.16 Necessary number of PCR cycles for different size of PCR targets. Three different sets of primers were used for the 35- and 45-cycle amplifications: (**1**) L29/H78, (**2**) L16179/16331, and (**3**) L16099/H16331. The same forensic sample was used for all the reactions except the positive DNA.

III.2.3.5 Reconstructive PCR

The polymerase extension technique or reconstructive PCR (Golenberg *et al.*, 1996) was tested on three archaeological samples. The method consists of two series of cycling reactions: a pre-treatment by a DNA *Taq* polymerase cycled extension followed by a standard PCR reaction. The rational behind the method is that the DNA polymerase in presence of dNTPs will repair nicks in the DNA and extend the DNA strands using overlapping fragments of DNA as template (by jumping PCR) and thus allowing the subsequent amplification of longer fragments. Thus, first PCR reactions included all standard PCR reagents *except* the primers. Five-µl of the aDNA solution were added and samples run in a Perkin Elmer 480 thermal cycler as follows: 94 °C for 11 min (enzyme activation), 94 °C for 30 sec. (denaturation), 50 °C for 30 sec. (annealing-like phase) and 72 °C for 3 min. (extension), for 20 cycles. These pretreated samples were

used as template for a standard amplification reaction, now including the specific primers. PCR reagents and conditions were as described previously except that the amplification reaction was run for 35 cycles and the primers used flank a 420 bp region of the mtDNA HVS-I (Ginther *et al.*, 1992). The tested archaeological samples had previously failed to amplify such PCR fragment. After reconstructive PCR, this longer HVS-I fragment could be amplified, yet it turned out to be a contaminant. The two-round approach probably increases the chance of contamination, as has also been described for standard second round PCRs (Kolman & Tuross, 2000). Still, the reconstructive PCR may be a suitable technique for non-human aDNA studies where contamination is less frequent.

III.2.4 DNA sequencing strategies

Positive PCR products were cleaned up by the QIAquick PCR purification kit and automatically sequenced (ABI Prism, Model 377, Version 3.0). All mtDNA products were first sequenced for the forward or light (L) strand using the same primer as for the amplification reaction. Only few fragments were reverse sequenced (see below), as sequencing of the heavy (H) strand of mtDNA does not yield good quality sequences due to its high purine (A and G) content (Parson *et al.*, 1998).

When a transition from thymine to cytosine at position 16189, along with a transversion from adenine to cytosine at positions 16182-3, were present an uninterrupted polycytosine tract occurred in region HVS-I (Bendall & Sykes, 1995) (see Fig. III.17). Direct sequencing of HVS-I sequences with these polymorphic variants produces a characteristic blurred DNA chromatogram for the nucleotides beyond the tract. This is due to slippage (i.e. erroneous annealing) of the two DNA strands being extended during the amplification and/or sequencing reaction. This was observed for the skeletal remains D18, D21, D23, G23 and M11 (and also putative relatives rel15, rel68, rel85, rel107 and rel116). In these cases, a smaller PCR fragment was amplified using the internal primer L16209 which anneals right at the cytosine tract and thus enables reading the sequence beyond this point (see position in Fig. III.5). These fragments were also sequenced for the reverse strand using primer H16331.

Figure III.17 DNA electropherogram showing sequence slippage beyond a polycytosine tract due to mutations at nucleotide positions 16183 (A→C) and 16189 (T→C).

III.2.5 Genetic identification

A summary of the genetic identification analyses will be presented here –for complete details see Appendix B.

DNA could be isolated from all the dental-powdered samples. Although some forensic specimens initially showed inhibition of the amplification reaction this could be overcome by further purification with InstaGene™. The *cold-step* and addition of BSA were always performed, as they improved the yield of the amplification. PCR products of both HVS-I and HVS-II, comprising a total of 610 bp, were finally obtained and sequenced for all 32 samples. The full length DNA sequences were compared to the human mitochondrial reference sequence (Anderson *et al.*, 1981) to determine the number and positions of variable sites and mitochondrial DNA haplotypes. Population frequency of the observed haplotype was evaluated based on a human mtDNA worldwide database (Handt *et al.*, 1998). Sequences are shown in Table III.4.

Among the 32 HVS-I sequences, 27 distinct haplotypes (ht) were observed. These could be further differentiated into 29 ht when HVS-II data was included (as samples [D8-D12] and [D23-M11] share HVS-I but not HVS-II polymorphic sites). Interestingly, a large proportion (40%) of the ht could be classified as Amerindian haplogroups (hg). The remaining sequences were assigned into European haplogroups. It should be noted that in addition to the European colonization of the Americas in the 15th century,

Argentina has had a large immigration from Europe during the last century (Devoto, 2003). This influence is expected to be more evident in urban populations as the 'sample' analysed. Haplotype diversity of the 'sample' was high (GD = 0.98 ± 0.03), as would be expected for a mixed population.

Sample	Polymorphic HVS-I sites								Polymorphic HVS-II sites					
Derqui cemetery														
D1[a]	147	172	223	248	320				152	199	204			+/+
D2	183	188	189	217	266									+/+
D3	129c	189							151	217				+/+
D4														2+/+
D5	256	270												+/+
D6	291	311												+/+
D7[b]	223	298	325	327					249d	190d	191d			2+/+
D8	311													
D9	291													2+/+
D10[c]	126	300							185	228	295			0/+
D11[b]	223	298	325	327					249d	190d	191d			2+/+
D12	311								152					0/+
D13[a]	147	172	223	248	320				152	199	204			+/+
D14	304g													+/+
D15	224	311							n/a					
D16	126	224							185	228	295			0/+
D17	218	261							152					+/+
D18	182	183	189	217	274				73	152				0/+
D19	189	223	298	325	327				185	249d	190d	191d		+/+
D20	129	185	189	193d	214	223	249	311	73	195	200			+/+
D21	183	189	217	291					151					2+/+
D22[c]	126	300							185	228	295			0/+
D23	183	189	217											2+/+
D24	162								73					+/+
D25	126	292	294	296	304				73	151	152			+/+
La Plata cemetery														
H13	223	311							73					+/+
G23	147	183	189	217					143	195				0/+
C26	223	242	311	325	327									+/+
Moreno cemetery														
M6	126	269	289	311										0/+
M11	183	189	217											0/+
Villegas cemetery														
V796	126	241							150	185	189	195	263	+/+
Avellaneda cemetery														
D8-42	111	223	290	319	362				73	146	150	235	263	+/+

Table III.4 Polymorphic positions along the HVS-I and HVS-II regions (total of 610 bp) for the 32 forensic samples. Numbering according to Anderson et al., 1981 (for HVS-I, prefix 16 was omitted for brevity). Mutations are all transitions unless stated. (d): deletion; (+): No. of insertions at the 309/310 cytosine tract of the HVS-II. (a), (b) and (c) indicate skeletal remains sharing the same mtDNA haplotype.

Among the 29 haplotypes observed, 26 were single haplotypes (observed only once) and 3 haplotypes were observed in three sets of two skeletal remains. In particular, haplotype (a) is shared between samples D1 and D13, haplotype (b) is shared between samples D7 and D11, and haplotype (c) is shared between samples D10 and D22 (see Table III.4), suggesting that these pairs of individuals might be maternally related to each other. However, it is not possible to determine the degree of relatedness of these individuals by mtDNA analyses alone. Samples D1, D7, D11 and D13 belong to young male bodies and samples D10 and D22 to females, the latter one to an older woman. Further nuclear DNA analyses of these two female samples allowed established the first-degree kinship (i.e. mother-daughter) between these two individuals (see Appendix B). Comparative analyses of the [D7-D11] pair and their putative relative yielded an interesting result. The mtDNA haplotype shared between these two individuals presents 9 differences with respect to the CRS (Anderson *et al.*, 1981), of which 3 are deletions and 2 are insertions. In addition, this haplotype is rare among mtDNAs present in a worldwide database (Handt *et al.*, 1998). When this mtDNA sequence was compared to the sequence of the presumed relative (rel112), they matched the entire 610 bp region except for one base. The mtDNA sequence obtained for rel112 presents one HVS-I variable site too many (a T→C transition at position 16189), which in theory would rule out the presumed relatedness. However, this position has been described to have a high substitution rate (Bendall & Sykes, 1995). Furthermore, differences in sequences between biological mother and child (observed as heteroplasmic positions in the child's sequence at this point) have been reported (Huhne *et al.*, 1999). These are generally seen as T/C heteroplasmic points (i.e. positions where a secondary peak is observed) in the child, whereas the corresponding sequence of the mother is homoplasmic (Huhne *et al.*, 1999; Lutz *et al.*, 2000). This might be the case for individual rel112, who might have gained a *de novo* C→T mutation at 16189. If so, the rest of the family would still present the wild-type base at this position, which would explain the mismatched sequences obtained. As the historical data strongly supports individual 112 as [D7-D11]'s aunt, it seems unreasonable to rule out this hypothesis based only on this result. It would therefore be crucial to analyse another relative (or other nuclear markers) to elucidate the identity of these skeletal remains.

When all samples analysed, i.e. skeletal remains and presumed relatives, were compared, mtDNA haplotypes belonging to fourteen skeletal remains matched those of living relatives (see Appendix B), suggesting maternal biological bond between them. In particular, the skeletal remains D4, D5, D6, D8, D16, D17, D19, D10 (and D22), H13, G23, C26, V796 and D8-42, matched the relatives 70, 93, 67, 59, 8 (and also 9), 77, 68, 92, 136, 149, 150, 100 and 36, respectively. In all the cases the sex and estimated age of the skeletal remains agreed with those of the missing persons. The observed mtDNA (except for D4) variants proved to be either unique or rare (<1%) among the mtDNAs present in a worldwide database (Handt *et al.*, 1998). This strongly confirmed the biological relationship between such individuals.

III.3 Conclusions

The thorough set-up of the methodology to extract and amplify DNA from poorly preserved hard tissues proved to be extremely necessary. Had this preliminary study not been undertaken, the subsequent archaeological study would have not been possible. DNA from forensic and ancient material behave in a similar manner due to the low number of templates present in the tissue. However, aDNA from forensic samples does not seem to be as damaged as that from archaeological material. Whereas replicates of forensic DNA sequences were always identical, archaeological samples yielded several types DNA artifacts (see Chapter IV).

In addition, the genetic analyses carried out in this preliminary study also allowed the identification of 14 victims of human rights' abuses. The remains have now been returned to their families for righteous burying –this will help them to overcome these distressing circumstances.

Chapter IV – Genetic diversity of mtDNA from ancient England

IV.1 Introduction

Analysis of DNA sequences in order to trace back population history and micro-evolutionary events has grown enormously in the past decade. The comparison of extinct populations with their extant counterparts permits the direct reconstruction of the evolutionary history of the population through time. Given that population genetics and molecular evolution are based on the study of genetic variability estimating the amount of variation is fundamental.

IV.1.1 Measurement and patterns of genetic variation

Genetic diversity is the variation in the genetic composition of individuals within or among species or populations. For diploid organisms diversity is usually measured as the number of heterozygous loci across individuals. For haploid organisms, or haploid markers such as mtDNA, genetic diversity can be measured at the gene and at the molecular level. Gene or haplotype diversity is simply defined in terms of gene frequency (Nei, 1987). Other estimators, mostly derived from ecological measures of species diversity, take into account both the richness and the evenness of genetic variants in the population (Hurlbert, 1971). Richness refers to the number of variants and evenness is defined as the ratio of the observed diversity to the maximum diversity, with this being when all the variants or haplotypes are equally abundant. In general, estimations of diversity are highly influenced by sample size (Nei, 1987). At the DNA level, diversity can be measured by calculating the average differences between pairs of sequences or the evolutionary distance between such sequences, assuming a particular model of molecular evolution. There are several methods for estimating evolutionary distances depending on the pattern of nucleotide substitution (see Nei, 1987). These distances vary with respect to the assumptions regarding the relative frequency of each

nucleotide and the transition/transversion bias in the sequence analysed. They can also be further corrected to account for non-homogenous substitution rate along the sites analysed (Wakeley, 1993; Meyer *et al.,* 1999), as is the case of mtDNA.

The amount and the pattern of molecular diversity depend in large part on the past demography and the size of a population. It is known that large populations maintain more diversity than small populations (Kimura & Crow, 1964) and that genetic drift is stronger in smaller populations, leading to a more rapid change in gene frequencies. Reduction in population size (bottleneck effect) is also known to reduce the amount of variability present (Nei, 1975). More recently, it has been realised that demographic expansions may leave a signature in the pattern of molecular diversity, such as phylogenies with star-like structures (Slatkin & Hudson, 1991) and unimodal distributions of pairwise differences (Rogers & Harpending, 1992; Harpending, 1994). In a case of population growth the resulting curve of mismatch distributions will travel at rate that depends on the mutation rate (μ) and the value of its crest will depend on this rate and the number of generations passed (Rogers & Harpending, 1992; Harpending *et al.*, 1993). Thus, the location of the wave provides a rough idea of the time (in units of $1/2\mu$) since the episode of population expansion.

The mtDNA analyses of European populations for mismatch distributions clearly did not fit the expected curves for stable populations, but instead matched the expectation for populations under expansion (e.g. Comas *et al.*, 1996; Pinto *et al.*, 1996; 1997; Varesi *et al.*, 2000). Accordingly, phylogenies are typically star-like, depicting recent shallow diversity (e.g. Bertranpetit & Cavalli-Sforza, 1991; CorteReal *et al.*, 1996; Richards *et al.*, 1996; 1998). In comparison with African and Asian populations (Chen *et al.*, 1995; Watson *et al.*, 1997), European populations generally exhibit lower levels of diversity and appear very similar to each other (Pult *et al.*, 1994; Torroni & Wallace, 1994; Sajantila *et al.*, 1995). Although substantial mtDNA diversity is present, most of the genetic difference is observed within rather than between populations so that very little differentiation is observed (Melton *et al.*, 1997) (see Chapter V). Nonetheless, there are exceptions, as some populations have proven to be genetically distinct. Examples of these are the Saami and the Basques.

The Saami – the origin of the Saami is obscure, as they are a highly fragmented population (Cavalli-Sforza *et al.*, 1994). They are found in the north of four different countries, Norway, Sweden, Finland and the former URSS. Physically they show considerable differences and vary from a dark to a blonde phenotype. They speak eight different languages, all belonging to the Uralic family, and yet are mutually intelligible (Ruhlen, 1987). Trees of genetic distances based on nuclear-coded genes show Saami as extreme outliers of Europe (Cavalli-Sforza *et al.*, 1994). At the mtDNA level, they are also highly differentiated from all the other European composition, despite showing very low haplotype diversity and medium nucleotide diversity. Their genetic differentiation is due to a few private haplotypes and, above all, a striking skewed distribution of relatively common haplotypes (i.e. haplotypes shared with other populations) (Sajantila *et al.*, 1995) (see Chapter V.2.2). Although admixture with other populations has been reported (Lahermo *et al.*, 1996), they have resisted assimilation and have lived mainly in genetic isolation, resulting in the genetic pattern now observed (Cavalli-Sforza *et al.*, 1994; Sajantila *et al.*, 1995).

The Basques – this population is today confined to a narrow zone in the Western Pyrinees, from the southwest of France to the northeast of Spain. They speak the Basque language, a non-Indo-European language of unknown origin. Some remote resemblance has been found, however, with North Caucasian languages (Gamkrelidze, 1990). It is believed that Basques were in the present area before the arrival of Neolithic people ~6,500 YBP, spoke a proto-Basque language and retained their distinctiveness in spite of the 'wave of advance' of the Neolithic migrants (Bertranpetit & Cavalli-Sforza, 1991). Renfrew (1987) proposed that the Indo-European languages had an origin in Anatolia, Turkey, and that their spread would be associated with the spread of farming during the Neolithic. This view would agree the pre-Neolithic origin of the Basques and the retention of their uniqueness due to lack of admixture with the Neolithic farmers (see Chapter V). Analyses of classical markers (e.g. blood groups) showed the distinctiveness of the Basques, supporting the interpretation of their origin as an ancient diversification in the place they occupy today. Mitochondrial DNA data, however, did not show clear differentiation between the Basques and other Europeans (Bertranpetit *et*

al., 1995). Basques have low pairwise differences between individuals, despite presenting high haplotype diversity. This is because the majority of their HVS-I sequences are restricted to the lineage of the CRS (Anderson *et al.*, 1981) and its derivates (Côrte-Real *et al.*, 1996). Also low numbers of steps in the most parsimonious tree links all the sequences. Many mutations are present in lineages that already presented a previous mutation, so even the most distant sequences can be connected by intermediate sequences which are present in the population. This pattern is consistent with a population found by few close lineages and probably, few individuals. This agrees with the model of a differentiation of the Basques *in situ* through drift and subsequent preservation of genetic features after a demographic expansion, and thus with the existing evidence that considers the Basques as a relict of the ancient population of Europe (Bertranpetit *et al.*, 1995).

IV.1.2 The British Isles

Genetic variation of the British Isles has been studied by analyses of blood groups (Bodmer, 1993; Mascie-Taylor & Lasker, 1996), serum proteins (Mastana & Sokol, 1998), HLA (Bodmer, 1993) and other classical genetic markers (Falsetti & Sokal, 1993). Little work has been done on mtDNA diversity which showed very low nucleotide diversity (1.1%) (Piercy *et al.*, 1993). Classical genetic markers showed zones of genetic change which have been interpreted as reflecting physical barriers as well as cultural and linguistic differences due to historical migrations and settlement patterns. In addition, comparative analyses of Y-chromosome haplogroups distribution between Wales, England and Frisia and between Wales, Orkney and Norway have suggested, respectively, a substantial male Anglo-Saxon and Viking migration into different parts of the British Isles (which can be still distinguished today) (Wilson *et al.*, 2001; Weale *et al.*, 2002). On the other hand, a female component has not been so far clearly identified possibly due to the more reduced structure in the mtDNA variation in Europe. So far, results would suggest either that females have played a minor role during these demic –or cultural– transitions or that their genetic fingerprint has been erased by other demographic events. Mitochondrial DNA analyses of the ancient English population can help elucidate this issue.

IV.1.3 Aims

This chapter will directly investigate the genetic variability of populations from Anglo-Saxon England. This will be carried out in the context of the variability presently observed in Europe and, in particular, in England. The genetic relationship between ancient and modern England will be analysed by comparisons of several estimators of haplotype and nucleotide diversity and by construction of phylogenetic networks. Given the difficulties of the aDNA work, authenticity of the data will be analysed in detail.

IV.2 RESULTS

IV.2.1 Retrieval and authentication of aDNA sequences

IV.2.1.1 Extraction and amplification of ancient DNA

Human skeletal remains exhumed from five archaeological sites located in south and north east England were studied. Dates of the sites cover a period from 4th to 11th century. Overall, a total of 156 individuals were analysed. Of these, 31 were from Buckland, 28 from Lavington, 15 from Leicester, 23 from Norton and 59 from Norwich. Two to three dental samples per individual were used, adding up to a total of 319 teeth (62 from Buckland, 36 from Leicester, 56 from Lavington, 47 from Norton and 118 from Norwich). Teeth were broken so that several fractions of powdered dental material were obtained from each tooth (see Methods II.2.5). The first fraction of material collected consisted of dental pulp or its remnants adhered to the pulp chamber. The subsequent fractions consisted of inner parts of the tooth, mainly dentin. Collection of enamel was always avoided, as it has been reported to inhibit the PCR reaction (Montiel *et al.*, 1997). Overall, a total of 721 fractions of powdered dental samples were subjected to the DNA extraction.

Retrieval of ancient DNA (aDNA) was assessed directly by PCR amplification of the extracted samples. 5 µl of the aDNA solutions were subjected to the optimised PCR

reaction described in Chapter III. A 264 bp fragment (including primer regions) of the mitochondrial DNA (mtDNA) control region was the PCR target.

Result of the amplification reaction, and therefore of the DNA extraction, was detected by agarose gel electrophoresis of the PCR products. Samples that failed to amplify were subjected to a second and a third PCR reaction, using the same initial aDNA extract solution as template. Second round PCRs (i.e. where first round PCR product is used as template for a second PCR) were not undertaken as they have proven more prone to contamination (Sullivan *et al.*, 1992; Allen *et al.*, 1998).

Performing a 'cold step' (i.e. leaving the DNA extraction solutions at 4 °C for 3 days before the amplification reaction; Montiel *et al.*, 1997) as well as adding BSA to the PCR mixes, facilitated the amplification reaction of aDNA samples (see Chapter III). Although the mechanism is not well understood yet, it is thought that the 'cold step' helps overcome inhibition by causing the inhibitory substances to adhere to the walls of the propylene tubes. In addition, BSA stabilises the DNA polymerase, enhancing the amplification reaction. As for the forensic samples, failure of amplification due to inhibitory compounds was often associated with absence of primer-dimers.

To remove inhibitory compounds, samples which repeatedly failed to amplify were subjected to a further purification step with the chelating resin Instagene™, as described in Chapter III. The resulting aDNA solution was used as template for a standard PCR reaction. Summary results of the DNA extractions and amplifications are shown in Table IV.1. Full details are given in Appendix C.

Site	indiv.	teeth	extract	+ amplif. (%)
Leicester	15	36	84	45 (53.6)
Norton	23	47	132	76 (57.6)
Buckland	31	62	103	53 (51.5)
Lavington	28	56	92	53 (57.6)
Norwich	59	118	310	157 (50.6)
Total	156	319	721	384 (53.3)

Table IV.1 Summary of extraction and amplification success rates for the five archaeological sites. (+ amplif.): positive amplification of the extract.

Overall, of the total 721 aDNA extracts, 384 (53.25%) yielded PCR products. However, of these, a rather large number yielded DNA sequences that were either

unreadable or from contaminating DNA. Therefore this value cannot be interpreted as a DNA extraction success rate. In any case, the values obtained for each archaeological site were on the whole very similar. Positive amplification ranged from 50.6% (Norwich) to 57.6% (Norton and Lavington) with $\bar{x} = 54.2 \pm 3.3$.

IV.2.1.2 Ancient DNA sequencing problems

PCR reactions showing a band of the correct size were purified and subjected to DNA sequencing using, in first instance, the same amplification primers. However, not every amplification product generated a DNA sequence. PCR products that showed a faint band in the agarose gel often produced very low quality template which resulted in poor or unreadable DNA sequence data (see Appendix C). This may be due to a loss of PCR product during the purification step, which, given the initial low quantity, resulted in an inadequate amount of DNA template for the sequencing reaction.

For the PCR products that did yield DNA sequences, it was noticed that forward sequencing was often clearer to read than reverse sequencing. This has been reported elsewhere (Parson *et al.*, 1998). The purine rich H-strand of the human mtDNA control region frequently gives more sequencing problems than the pyrimidine rich L-strand. Therefore, save for specific cases (see below), only forward sequencing was carried out.

Still, some of the L-strand DNA sequences presented background noise. Instead of a neat dye peak per nucleotide, the electropherograms showed several small overlapping and messy peaks as background (see Fig. IV.1). When this noise was too high the sequences became difficult to read and therefore unreliable (shown as *unread.* in Table IV.2). This result has been previously observed for sequencing reactions of ancient DNA samples (Montiel, pers. comm.).

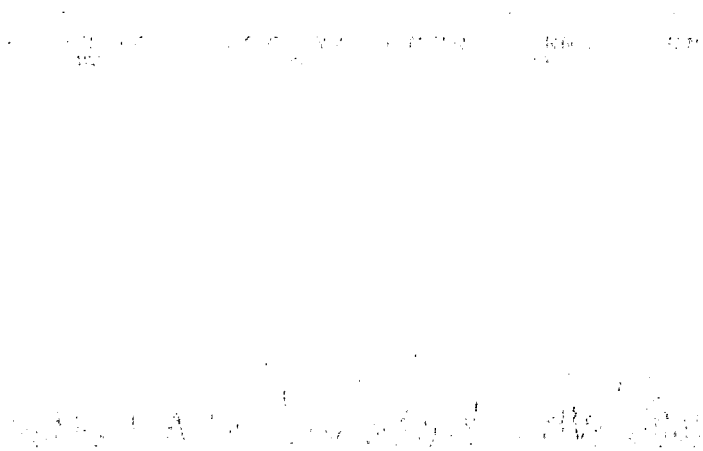


Figure IV.1 DNA electropherogram showing a sequence with high background noise.

As described in Chapter III for the forensic samples, the presence of a transition T→C at position 16189 and transversion A→C at position 16182-3 (observed in samples R3 and w14) rendered an uninterrupted polycytosine tract, producing sequence slippage. In order to determine these sequences, a smaller PCR fragment was amplified using the internal primer L16209 (see Chapter III.2.4). Moreover, to further corroborate such samples, these PCR products were also sequenced for the reverse H-strand.

IV.2.1.3 Identification of contaminating sequences

To monitor contamination, negative controls were carried out alongside the DNA extractions and PCR amplifications. DNA extractions were carried out in sets of 12 samples: 10 dental samples and 2 extraction controls (one with no source of DNA and the other with non-human DNA; see Methods II.2.4.3) to check for presence of contaminants in the reagents or during the extraction procedure. PCR amplifications included the ancient DNA extraction solutions, the two extraction controls and a PCR control. The PCR control consisted of all the PCR reagents and water instead of DNA. If either of these controls yielded amplification, the resulting PCR product was purified and sequenced in order to identify the source of contamination. Several mtDNA types were repeatedly found as contaminating DNA. These were the Cambridge Reference sequence (CRS) (Anderson *et al.*, 1981) and sequences with variable nucleotides at positions [16126], [16145-16176-16223], [16126-16261-16274] and [16175-16213], where the numbers refer to the positions at which the sequence differs from the CRS.

Being the most likely source of contamination, DNA of lab members was sequenced for comparisons. In this manner, it could be determined that the three latter sequences correspond to my own mtDNA sequence and two other researchers' (who also work in the aDNA lab) sequences. The origin of the [CRS] and [16126] contamination is still unknown, but it might be due to remaining contamination in disposable material, as was found previously (Schmidt *et al.*, 1995). The [CRS] and [16126] are two of the most common European mtDNA haplotypes. Details of identified contaminated sequences are shown in Appendix C.

PCR controls were usually clean, with the extraction blanks being more prone to contamination. However, clean extraction and amplification controls did not totally rule out contamination in the tested samples. It was noticed that even when controls were negative some sequences turned out to be contaminated (usually of the same type as mentioned above). This type of contamination was more common than contamination in extraction blanks and can be explained by a carry-over or enhancing effect of the ancient DNA extract (Kolman & Tuross, 2000). A summary of the amplification results now including the contamination rate and the non-readable sequences is shown in Table IV.2. The proportion of unreadable sequences ranged from 9.8 (Lavington) to 10.7 (Leicester and Buckland). As would be expected, the origin of the samples had no bearing on this type of artifact. The proportion of positive amplifications which turned out to be the result of a contaminating DNA, ranged from 9.7 (Buckland) to 16.5 (Norwich), indicating that contamination was not site-specific either.

Site	extract	+ amplif. (%)	unread. (%)	cont.? (%)
Leicester	84	45 (53.6)	9 (10.7)	11 (13.1)
Norton	132	76 (57.6)	14 (10.6)	15 (11.4)
Buckland	103	53 (51.5)	11 (10.7)	10 (9.7)
Lavington	92	53 (57.6)	9 (9.8)	12 (13)
Norwich	310	157 (50.6)	33 (10.6)	51 (16.5)
Total	721	384 (53.3)	99 (10.5)	76 (13.7)

Table IV.2 Proportion of PCR products which turned out to be non-utilizable data, i.e. either unreadable or contaminated DNA sequences.

IV.2.1.4 Ancient DNA artifacts

Among the aDNA sequences obtained a number of chimeric molecules were observed. This was noticed when different amplifications of the same sample yielded similar yet not identical sequences. Sample Nw18 is an example of this: extract w18.1.1 yielded a sequence with variable sites at positions 16126, 16163, 16186, 16189 and 16294. A different extract from the same sample (extract w18.2.1) yielded a sequence which perfectly matched the previous one, corroborating it. However, extract w18.2.2 presented the same variable sites *plus* positions 16261, 16262 and 16270. These three positions are thought to be misincorporations rather than actual variable sites.

Sample Nw1 is an extreme example of these 'artifacts'. Its DNA extract w1.1.1 yielded a sequence with variable sites at positions 16270 and 16311. Extract w1.2.1 however, showed another 10 variable sites, namely 16128, 16142, 16147, 16223, 16232, 16264, 16268, 16270, 16311 and 16327. All these substitutions, save the probably real site 16311, were C→T transitions. A similar case was observed for sample L11 for which the sequence with variable sites at positions 16256 and 16270 was confirmed by triplication. However, another extract of the same sample (L11.1.2) presented another 8 variable sites at positions 16128, 16218, 16228, 16244, 16248, 16278, 16292, 16303. Again, all these were C→T transitions. Other extracts produced similar situations.

Another type of artifact observed were sequences with a mixture of two distinct nucleotides at a given position, coded as N by the software program. Unlike the electropherograms showing background noise described previously, in this case, these positions presented clearly superimposed peaks, where the height of the second peak was 50% or more of the main peak (see Fig. IV.2). This phenomenon is known as heteroplasmy.

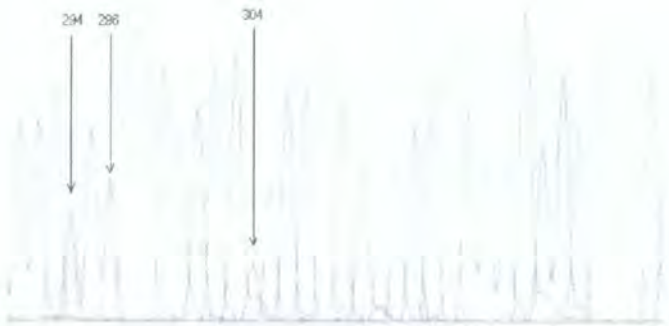


Figure IV.2 DNA electropherogram showing a sequence with 3 T/C heteroplasmic peaks at positions 16294, 16296 and 16304 (prefix 16 is omitted). The software labels the nucleotide position as N.

In general, a heteroplasmic position is the result of amplification and sequencing of a mixture of DNA templates. If both templates are present in similar concentration, after amplification they will be equally represented as PCR products. They will also be equally represented in the sequencing reaction mixture and as a result, in the DNA sequence, giving the characteristic superimposed peaks. The different templates can represent: a) an individual harbouring a mixture of wild type and mutant mtDNA (e.g. responsible for a disease); b) a mixture of DNA templates from different sources (i.e. exogenous contaminating DNA); and c) a damaged –endogenous– DNA template. In this study, heteroplasmic-like sequences are thought to be caused by the latter two situations.

Examples of heteroplasmy due to mixture with exogenous DNA are the sequences showing heteroplasmy at positions 16145, 16176 and 16223. As it was mentioned, these are the variable sites which identify my own HVS-I mtDNA sequence, therefore indicating contamination with traces of DNA probably from my skin cells. Another situation observed was the mixture of a (possible) ancient DNA sequence and other common mtDNA haplotype, as the CRS (Anderson *et al.*, 1981) or [16126], which, as above mentioned could be the result of contaminating traces in plastic-ware or others.

Heteroplasmy due to DNA damage can be exemplified with sample N19. Extract N19.1.1 presented positions 16192, 16256, 16270 and 16311 as variable sites. Another

extract (N19.2.1) confirmed this sequence. However, extract N19.1.2 and extract N19.2.4 showed the same variable site and two heteroplasmic positions at 16290 and 16292 and at 16184 and 16185, respectively. All these heteroplasmic positions are mixtures of cytosine and thymine. As for the aDNA artifacts previously described, this high rate of C→T transition would be an indicator of damaged DNA templates. This phenomenon has also been observed in other ancient DNA studies (Kolman & Tuross, 2000; Merriweather, 2000; Montiel *et al.*, 2001).

Certainly, none of these artifact sequences were included in the population analyses, yet they were valuable results as this behaviour is typical of aDNA material and therefore suggested that such samples did contain ancient DNA. Thus, the for the calculation of the aDNA extraction success rate, not only the extracts that yielded authenticated DNA sequences (see below) but also those extracts that were believed to contain aDNA were included. The aDNA extraction success rate ranged from 20.7% (Lavington) to 25.8% (Norton) with a $\bar{x} = 21.2 \pm 2.8$ (Table IV.3). The overall success rate was 20.7% which although low, is well within the range observed for ancient DNA (e.g. Merriwether *et al.*, 1994).

Site	indiv.	teeth	extracts	aDNA extr. (%)
Leicester	15	36	84	18 (21.4)
Norton	23	47	132	34 (25.8)
Buckland	31	62	103	20 (19.4)
Lavington	28	56	92	19 (20.7)
Norwich	59	118	310	58 (18.7)
Total	156	319	721	149 (20.7)

Table IV.3 Extraction success rate for the ancient samples, including samples that were either replicated or showed characteristic aDNA behaviour.

IV.2.1.5 Replication of DNA sequences

To validate the aDNA data, sequencing of each sample was repeated several times. A sequence was considered to be authentic when *different* extracts from *different* dental samples (of the same skeletal remain) yielded identical sequences. Only DNA sequences which were corroborated by replication were included in the population analysis.

Moreover, to further corroborate the reliability of the data obtained, three samples were analysed in two independent labs. Samples N15 and Nw38 were analysed at the Smithsonian National Museum of Natural History, Washington DC, USA and sample Nw49 was analysed at the Ancient Biomolecules Centre, Department of Zoology and Biological Anthropology, University of Oxford, UK. Samples N15 and Nw38 were amplified using internal primers, resulting in two smaller PCR products. For both samples, the first half of the sequence perfectly matched the respective sequence obtained in this study, whereas the second half was found to be a contamination product. In any case, the majority of the variable sites present in each sequence were confirmed, that is 2 of the 3 variable sites for sample N15, and 5 of the 6 variable sites for sample w38. The full length HVS-I mtDNA sequence of Nw49 obtained in the second lab totally matched the one obtained in this study. These results obtained by independent labs are strong support for the authenticity of the aDNA data obtained in this study.

IV.2.2 General description of ancient mtDNA sequences

IV.2.2.1 Variable sites

For the five archaeological sites, a total of 48 authenticated aDNA sequences were obtained. Of these, 17 were from Norwich, 12 from Norton, 6 from Lavington, 7 from Buckland and 6 from Leicester. The sequence included 207 bp from nucleotide position (np) 16123 to 16329, numbered according to Anderson et al. (1981). For simplicity, hereinafter 16000 will be assumed and positions will be numbered from 123 to 329. Transversions, as well as deletions and insertions, will be further indicated with a suffix. Sequences are shown in Figure IV.3.

```
                1111111111111222222222222222222233333
                2256678888899011222235677799999900122
                6932323467923778123496001481234681410 4

CRS             TGGAATACCCTCCATCCCCTCCCCTGCCCACCTCTTCT          R  N  B  L  W

L2/N9/w55       .......................................         1     1  1
L20/w27/w37     C......................................                1  2
N12             C.........................C........              1
N1              C........T.....................TT.....C          1
B10/B9/R8       C...........................TT..C...         1     2
w49             C..........T.G..............T...C...                      1
w29             C..........T..........T.......T...C...                   1
w18/w32         C...G...T.C.....................T.......                2
w38             C...G...T.C.-...................T.......                1
w25             C...G...T...-.......T..........T.......                  1
N15             C.A............................T.......              1
N1              .A.....................................              1
N5              .A.........T.......................C..              1
N16             .A................T.....................              1
w22/R13         ...G...................................         1           1
B22/B31         ...G.........................T.........                2
w5              .....C...TC...C...T....................                  1
B14             ..........C............................                1
w30/R7          ..........C............T..............         1           1
w14             .......C...C..........T................                  1
R3              .......C...C........T.......T..........         1
w6              ..........A........T.....C.T...........                  1
N20             ........T..........T..........T.........              1
N57             ....................T..........T.........              1
B11             ...........T......T......A.......T....                 1
L3              ...........T.........T.T................                  1
N19/N7          ...........T.........T.T...........C..              2
N11             ...........T.........T.T...T...........                  1
L10/L11         .......................T.T..............                  2
R9              ................T....A...........C......         1
w51             ..................T.....................                     1
w21             ...................C....................                     1
w9              ....................C..........G.....C..                     1
B17             .....................................C..                 1
N3              .................................C..C..              1
R2              ................................T...C.T.         1
                                                                 ------------
                                                                 6 12  7  6 17
```

Figure IV.3 HVS-I mtDNA sequences of 48 ancient samples. Abbreviations for the archaeological sites: (B): Buckland, (L): Lavington, (N): Norton, (w): Norwich and (R): Leicester. The notation of the CRS (minus 16000) (Anderson *et al.*, 1981) is used for numbering the nucleotide positions (on top). Dots indicate identity with the bases in the CRS (first raw), a dash denotes deletion of the base, A, C, T and G are adenine, cytosine thymine and guanine, respectively. Transversion are shown in bold type. The column at the right hand side shows the distribution of haplotypes in the different settlements (same codes as above were used).

Among the 48 sequences there were 38 polymorphic sites (18.3%). At 37 substitution sites there were 40 substitutions and one indel (at position 193, shown with a dash (-) in Fig. IV.3). With the exception of 4 transversions (A→C at np 183, T→A at np 189, C→A at np 239 and A→C at np 293, shown in bold type in Fig. IV.3) all the remaining 36 substitutions were transitions. Most of the transitions concerned pyrimidines (29 out of 36) and the pyrimidine:purine ratio was 30:8. The transition:transversion ratio (*s:v* ratio) was 9:1 (36:4) a little higher than the one obtained by Tamura and Nei (1993), *s:v* ratio = 15:1. However, Tamura's value was calculated using the entire control region. For the 207 bp analysed in this study, a 9:1 *s:v* ratio is within the wide range (2:1 to 23:1) observed for the modern populations analysed.

Given that polymorphisms are defined in relation to the CRS, the HVS-I mtDNA sequence with the lowest number of polymorphic sites was the sequence identical to the reference sequence, observed in three individuals (L2, w55 and N9). The highest number of variable sites present in a single HVS-I haplotype was 6 (samples Nw38 and Nw25). The distribution of the number of polymorphic sites per sequence is shown in Figure IV.4

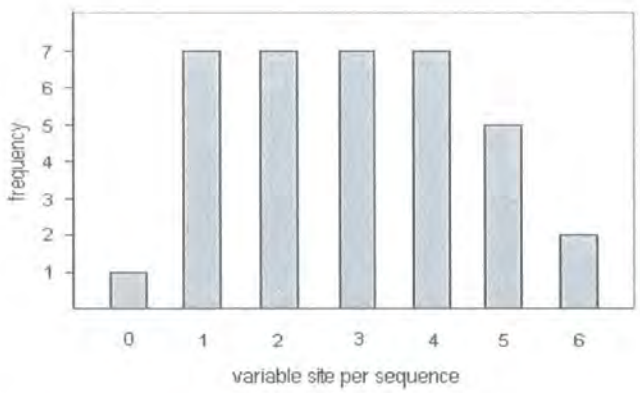


Figure IV.4 Distribution of number of polymorphic sites per sequence for the 48 ancient samples.

The two most frequent polymorphic sites were 126 and 294, with a T→C and a C→T transition, respectively, both found in 10 haplotypes. Positions 189, 192 and 223 were also very frequent, observed in 9, 7 and 7 haplotypes, respectively. Several single mutations (i.e. mutations occurring in only one sequence) were observed (41%). The observed frequency of mutation per site, along the 207 nucleotides is shown in Figure IV.5, which overall agreed with the data published by Meyer et al. (1999).

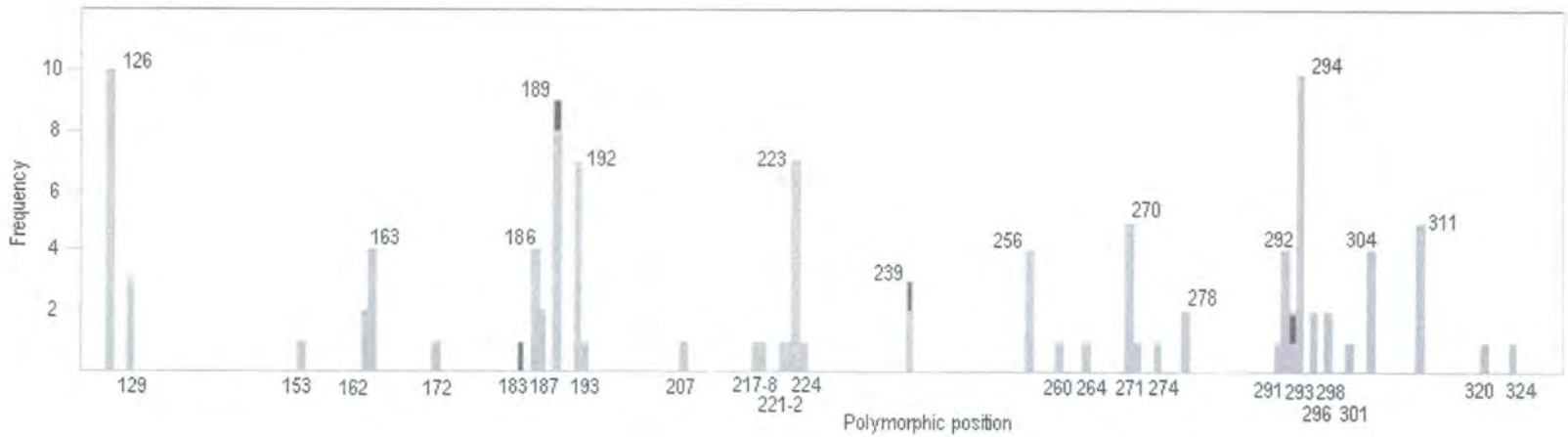


Figure IV.5 Distribution of polymorphic sites along the 207 bp of the HVS-I analysed. Numbering according to Anderson *et al.* (1981) (minus 16,000 for brevity). Darker sections indicate transversions.

IV.2.2.2 Haplotype distribution

Of the 36 different HVS-I haplotypes, 9 (25%) were multiple haplotypes (i.e. haplotypes found in more than one individual) whereas the remaining 27 (75%) were unique haplotypes (i.e. observed only in one individual among the 48 samples). The number of unique haplotypes was within the range observed for the modern populations ($\bar{x} = 72.7 \pm 10.97$).

Population	N	k	k/N	multiple	unique (%)
Leicester	6	6	1	0	6 (100)
Norton	12	11	0.92	1	10 (91)
Buckland	7	5	0.71	2	3 (60)
Lavington	6	5	0.83	1	4 (80)
Norwich	17	15	0.88	2	13 (87)
Ancient Britain	48	36	0.75	9	27 (75)

Table IV.4 Haplotype distribution for all ancient samples and for each archaeological site. (N): sample size, (k): number of haplotypes.

Five of the haplotypes (ht) were shared among individuals belonging to different settlements and the other 4 among individuals from the same settlement (see Fig. IV.3). Individuals L2, N9 and w55 presented the CRS haplotype, individuals L20, w27 and w37 shared ht [126], individuals B9, B10 and R8 shared ht [126-294-296-304], individuals R7 and w30 shared ht [189-270] and individuals R13 and w22 shared ht [162]. On the other hand, two individuals from Lavington (L10 and L11) shared ht [256-270], two individuals from Norton (N19 and N7) shared ht [192-256-270-311], two individuals from Norwich (w18 and w32) shared ht [126-163-186-189-294] and two individuals from Buckland (B31 and B22) shared ht [162-292]. This suggests possible biological relationship between these pairs of individuals, in particular, for those sharing haplotypes that are rare such as [162-292] (see Chapter V). Location of their graves is indicated in the respective maps of the cemeteries (see Chapter II.1.1.3). The two individuals from Norwich were buried at very close distance to each other, supporting their possible relationship. However, the samples from Buckland (which are very likely maternally related given their rare mtDNA) were very distant, suggesting that family members might not always be buried together (Parfitt, pers. comm.)

IV.2.3 Genetic diversity

Intra-population genetic diversity was estimated using several parameters, both at the haplotype and nucleotide level. At the haplotypic level, Nei's estimator of gene diversity (*h*, here stated as GD) (Nei, 1987), the Shannon index, its standardized version (Magurran, 1988) and the allelic richness were calculated. The two latter account for differences in sample size. At the nucleotide level, nucleotide diversity was calculated as the mean pairwise differences within populations (i.e. average of differences between pairs of sequences from each population) (Tajima, 1983), following Tamura and Nei's model (Tamura & Nei, 1993) with $\alpha = 0.26$ (Meyer *et al.*, 1999). Results are shown in Table IV.5

k/N, the ratio of haplotypes per sample size was calculated as a rough estimator of the diversity of the populations. It widely ranged from 0.17 to 0.98 with a $\bar{x} = 0.56 \pm 0.18$. As expected, the number of haplotypes increased with sample size, although very large sample size did not show a proportional increment (see Fig. IV.6).

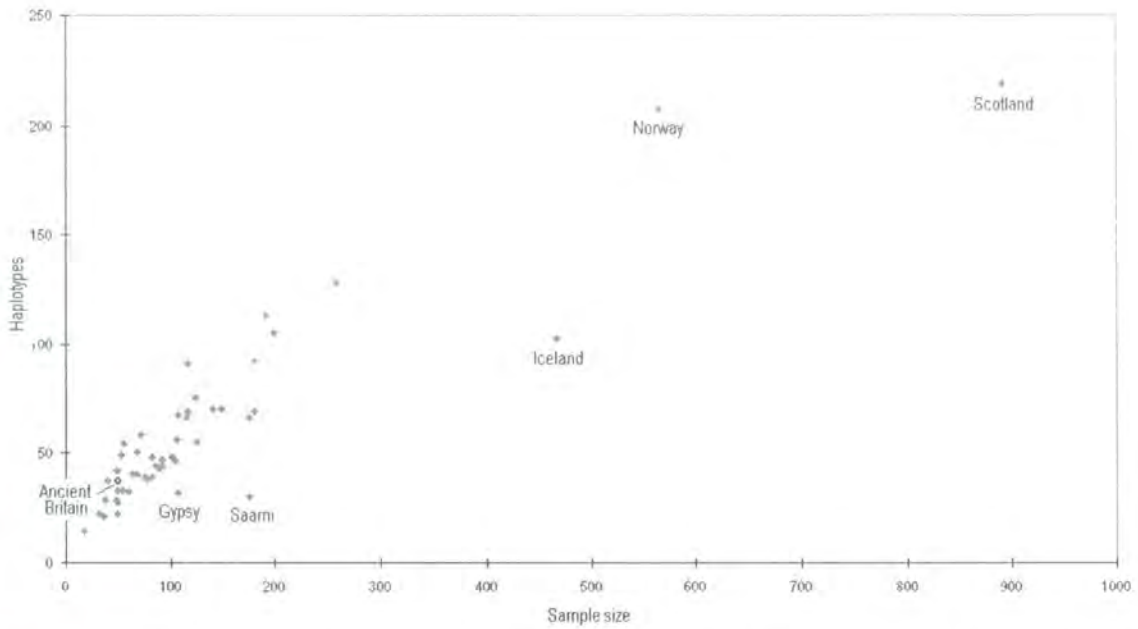


Figure IV.6 Number of haplotypes vs. sample size. The sample from Ancient Britain is indicated. Samples showing a comparatively small number of haplotypes per sample size are also indicated.

Population	N	k	k/N	GD ± sd	H_s	H'_s	$r_{(48)}$	$r_{(18)}$	π ± sd
Armenia	191	113	0.59	0.98 ± 0.004	-4.42	0.84	38.68	16.15	5.30 ± 2.57
Austria	116	66	0.57	0.95 ± 0.015	-3.69	0.78	33.04	14.24	4.33 ± 2.15
Basques	105	46	0.44	0.90 ± 0.024	-3.14	0.67	26.35	12.23	2.82 ± 1.50
Belarus	55	54	**0.98**	**0.99 ± 0.003**	-3.98	**0.99**	**47.24**	**17.90**	5.23 ± 2.57
Bulgaria	141	70	0.50	0.94 ± 0.012	-3.65	0.74	30.37	13.74	3.94 ± 1.98
Czeck Republic	83	48	0.58	0.97 ± 0.009	-3.59	0.81	32.48	14.88	4.17 ± 2.09
Denmark	48	28	0.58	0.92 ± 0.013	-2.92	0.76	28.00	12.86	4.85 ± 2.40
England:	344	154	0.45	0.93 ± 0.011	-4.06	0.70	31.44	13.59	3.81 ± 1.92
Cornwall	86	44	0.51	0.89 ± 0.030	-3.11	0.70	27.90	12.33	3.30 ± 1.71
central & north	258	128	0.50	0.94 ± 0.010	-4.05	0.73	32.46	14.01	3.98 ± 2.00
Estonia	149	70	0.47	0.94 ± 0.011	-3.67	0.73	30.55	13.81	4.10 ± 2.05
Finland	176	66	0.38	0.94 ± 0.012	-3.60	0.70	29.60	13.55	3.83 ± 1.93
France	64	40	0.63	0.94 ± 0.020	-3.33	0.80	32.06	14.11	3.70 ± 1.89
Georgia	124	75	0.60	0.96 ± 0.011	-3.85	0.80	34.05	14.68	4.68 ± 2.31
Germany:	473	211	0.45	0.96 ± 0.007	-4.50	0.73	34.75	14.70	3.96 ± 1.90
south	200	105	0.53	0.92 ± 0.016	-3.83	0.72	32.01	13.46	3.48 ± 1.78
north	107	67	0.63	0.98 ± 0.005	-3.96	0.85	36.20	15.81	5.03 ± 2.46
central	117	69	0.59	0.92 ± 0.020	-3.60	0.76	32.75	13.63	3.69 ± 1.88
Bavaria	49	33	0.67	0.96 ± 0.015	-3.28	0.84	32.49	14.70	3.94 ± 2.00
Greece	125	55	0.44	0.95 ± 0.008	-3.54	0.73	28.87	13.85	3.69 ± 1.88
Iceland	467	102	0.22	0.96 ± 0.004	-3.82	0.62	28.39	13.87	4.54 ± 2.23
Ireland	101	48	0.48	0.92 ± 0.020	-3.23	0.70	27.36	12.57	3.29 ± 1.70
Italy:	324	178	0.55	0.94 ± 0.010	-4.38	0.76	34.96	14.46	4.73 ± 2.32
North Italy	68	50	0.74	0.98 ± 0.007	-3.87	0.92	40.07	16.65	5.78 ± 2.80
Tuscany	49	37	0.76	0.96 ± 0.018	-3.36	0.86	36.33	15.08	4.94 ± 2.44
Sardinia	69	40	0.58	0.92 ± 0.027	-3.18	0.75	30.04	13.18	4.29 ± 2.15
Sicily	90	43	0.48	0.86 ± 0.035	-2.95	0.66	26.48	11.63	3.36 ± 1.74
Rome	48	42	0.88	**0.99 ± 0.007**	-3.67	0.95	42.00	16.95	5.68 ± 2.77
Karelia	83	39	0.47	0.92 ± 0.022	-3.15	0.71	26.97	12.78	3.57 ± 1.83
Norway	565	207	0.37	0.93 ± 0.009	-4.19	0.66	31.38	13.50	3.72 ± 1.88
Palestine	117	91	0.78	0.98 ± 0.005	-4.31	0.90	41.12	16.60	**5.83 ± 2.80**
Poland	37	21	0.57	0.94 ± 0.023	-2.77	0.77	n/a	12.50	**2.76 ± 1.49**
Portugal	54	33	0.61	0.91 ± 0.034	-3.00	0.75	29.96	12.95	3.49 ± 1.80
Romania	92	47	0.51	0.96 ± 0.009	-3.50	0.77	30.13	14.27	4.13 ± 2.07
Romano-Gypsy	130	36	0.28	0.81 ± 0.032	-2.64	0.54	18.49	9.76	5.39 ± 2.60
Russia:	287	151	0.53	0.98 ± 0.003	**-4.60**	0.81	38.30	16.15	4.98 ± 2.43
Adygei	50	27	0.54	0.93 ± 0.021	-2.92	0.75	26.20	12.54	4.54 ± 2.27
Uralic region	53	49	0.92	0.99 ± 0.005	-3.86	0.97	44.65	17.48	4.85 ± 2.40
north Ossetia	106	56	0.53	0.97 ± 0.005	-3.79	0.81	33.50	15.39	5.37 ± 2.61
north Eur.	40	37	0.93	**0.99 ± 0.006**	-3.58	0.97	n/a	17.41	5.19 ± 2.56
west siberia	38	28	0.74	0.98 ± 0.015	-3.22	0.89	n/a	15.40	3.82 ± 1.96
Saami	176	30	**0.17**	**0.80 ± 0.019**	**-2.17**	**0.42**	**13.23**	**7.41**	3.89 ± 1.96
Scotland:	1199	268	0.22	0.95 ± 0.039	-4.38	0.62	31.91	14.22	4.14 ± 2.06
mainland	891	219	0.25	0.95 ± 0.004	-4.25	0.63	31.25	14.00	4.09 ± 2.04
Western Is.	181	69	0.38	0.96 ± 0.070	-3.74	0.72	30.19	14.25	4.17 ± 2.08
Is. Skye	49	22	0.45	0.92 ± 0.020	-2.82	0.73	22.44	11.82	4.39 ± 2.20
Orkney	78	37	0.47	0.95 ± 0.014	-3.24	0.74	27.35	13.37	4.17 ± 2.09
Spain	181	92	0.51	0.93 ± 0.014	-3.82	0.74	32.24	13.83	3.75 ± 1.90
Sweeden	60	32	0.53	0.93 ± 0.020	-3.07	0.75	27.24	13.00	4.25 ± 2.14
Switzerland	76	39	0.51	0.96 ± 0.011	-3.36	0.78	29.13	14.01	3.37 ± 1.75
Syria	49	42	0.86	0.98 ± 0.010	-3.62	0.93	41.20	16.48	5.12 ± 2.52
Turkey	72	58	0.81	**0.99 ± 0.004**	-3.97	0.93	41.10	16.88	5.56 ± 2.70
Ukraine	18	14	0.78	0.96 ± 0.029	-2.55	0.88	n/a	14.00	4.43 ± 2.29
Volga-finns	32	22	0.69	0.96 ± 0.016	-2.96	0.85	n/a	14.27	4.39 ± 2.22
Wales	92	44	0.48	0.92 ± 0.020	-3.20	0.71	27.32	12.69	3.27 ± 1.70
Ancient Britain	48	36	0.75	0.98 ± 0.071	-3.49	0.90	36.00	16.11	5.51 ± 2.69

Table IV.5 Genetic diversity for the samples from ancient Britain and 50 European, Near Eastern and Caucasian populations. (N) sample size, (k) number of haplotypes, (GD ± sd.) gene diversity ± standard deviation, (H_s) Shannon diversity index, (H'_s) standardized Shannon index (H_s/H_s max). Haplotypic richness ($r_{(g)}$) based on a sample of g = 48 and g = 18, (π ± sd) nucleotide diversity (in %) ± standard deviation. Maximum and minimum values are indicted in bold type.

The ancient samples showed a k/N = 0.75 ranking among the first top ten populations, together with populations of East Europe and West Asia (European Russia, Belarus, Ukraine and Uralic region of Russia), Near East (Palestine, Turkey and Syria) and two Italian populations (Rome and Tuscany). However, modern England was at the other end of the k/N range with a k/N = 0.45. The Irish, Welsh and Cornish populations –thought to represent the native Britons– also showed low values (0.48, 0.48 and 0.51, respectively). However, Norway, who may represent the Scandinavian source population, was also within this range.

Gene diversity, Nei's estimator of the haplotype diversity showed a $\bar{x} = 0.941 \pm 0.043$, ranging from 0.80 to 0.99. Only small differences among the populations were observed. In any case, the ancient samples showed GD = 0.98 ± 0.07, ranking again among the first top ten populations (including East Europe and Russia, Near East and Italy). On the other hand, the modern English sample had GD = 0.93 ± 0.01, significatively lower than ancient England ($P = 1.6\ 10^{-6}$). If only central and north England were considered (i.e. excluding Cornwall, GD = 0.89 ± 0.03) a higher GD value was observed (GD = 0.94 ± 0.01) yet still significatively lower than the GD value obtained for ancient England ($P = 1.6\ 10^{-6}$ and $P = 8\ 10^{-5}$, respectively).

The Shannon index is usually used in Ecology to estimate species diversity and it has also been used in genetic analyses to estimate genetic diversity (Bertranpetit *et al.*, 1995; Comas *et al.*, 1997). However, correcting for sample size requires the use of the standardized version, H'_S (Magurran, 1988), which is simply the division of H_S by its highest possible value for a given sample size. Thus, a value of $H'_S = 1$ would imply that all the individuals in a population show different haplotypes. For the populations analysed H'_S ranged from 0.42 to 0.99, with $\bar{x} = 0.77 \pm 0.11$. The ancient samples ranked among the top ten populations with an H'_S of 0.90, similar to populations of the Near East, West Asia and North Italy. Again, modern England showed a much lower value of $H'_S = 0.70$. The Irish, Welsh and Cornish populations showed similar values of H'_S (0.70, 0.71 and 0.70, respectively). However, Germany (0.73) and Norway (0.66) also yielded low H'_S values.

The allelic (haplotypic, in this case) richness ($r_{(g)}$) another estimator which also accounts for differences in sample size was calculated. The standardisation in this case

is done by a method called rarefaction (Sanders, 1968; Hurlbert, 1971; Simberloff, 1972). Based on the larger sample set, rarefaction calculates how many species would have been found in a smaller sample of a specified number of individuals (g). Calculations were carried out for two specified sample sizes: g = 48 (the sample size of the ancient English sample) and also g = 18 (the smallest sample size among the 50 populations). Results are shown in Table IV.5. The haplotypic richness (g = 48) ranged from 47.24 to 13.23 ($\bar{x} = 31.93 \pm 6.16$). $r_{(48)}$ was not calculated for five populations as their sample sizes were smaller than 48 individuals (Poland, north European Russia, west Siberia, Ukraine, and the Volga-Finns). The ancient sample showed, by definition, a $r_{(48)} = 36$ which ranked among the top ten populations. Other populations with high values of allelic richness were Belarus, showing the highest value (47.24), followed by the Uralic region of Russia, populations of the Near East and some Italian populations. Modern England showed a medium value of 31.44. The population showing the lowest value was the Saami (13.23) followed by the Romano-Gypsies (17.5). Haplotypic richness calculated for g = 18 ranged from 17.9 to 7.41 with $\bar{x} = 14.14 \pm 1.87$. The ancient sample showed a value of 16.18 which was again among the top ten populations. Almost all of the populations were ranked in the same position as when a sample size of g = 48 was used for the calculation.

Regarding the nucleotide diversity, π ranged from 2.76 to 5.83, with $\bar{x} = 4.29 \pm 0.76$. The ancient English samples were among the 5 top populations with $\pi = 5.51$. Two Near Eastern populations (Palestine and Turkey) and two Italian populations (Rome and North Italy) were the other four populations showing the highest values. Modern England's nucleotide diversity was 37 places down in the list with $\pi = 3.81$. The population showing the lowest value was Poland (2.76) followed by the Basques (2.82), and the Celtic populations of Wales and Ireland (3.27 and 3.29, respectively).

The high levels of genetic diversity observed could be explained by an effect of the subdivision among the archaeological sites. To assess this hypothesis, internal genetic diversity was calculated for each archaeological site separately and also grouped according to their historical period (as described in Methods II.3.1). In addition, samples from modern populations which might have been source populations of the modern English sample were pooled together and the gene diversity estimated as above. This

made up sample, (referred to as *founder*) comprised the modern populations which are thought to represent both the native and the invaders of the British Isles. Representing the former, Wales (and to some extent Cornwall), and representing the latter Germany, Denmark and Norway (as possible Anglo Saxon and Vikings populations) were included.

Population	N	k	k/N	GD ± sd	H_s	H'_s	$r_{(18)}$	π ± sd
Leicester	6	6	1.00	1.00 ± 0.09	-1.79	1.00	n/a	6.08 ± 3.37
Norton	12	11	0.92	0.98 ± 0.04	-2.37	0.95	n/a	5.69 ± 2.93
Buckland	7	5	0.71	0.91 ± 0.01	-1.75	0.90	n/a	5.27 ± 2.89
Lavington	6	5	0.83	0.93 ± 0.01	-1.56	0.87	n/a	2.37 ± 1.49
Norwich	17	15	0.88	0.98 ± 0.02	-2.67	0.94	16.00	5.79 ± 2.91
Period								
Romano-British	6	6	1.00	1.00 ± 0.09	-1.79	1.00	n/a	6.08 ± 3.37
Early Saxon	25	20	0.80	0.93 ± 0.01	-2.94	0.91	15.45	4.88 ± 2.46
Late Saxon	17	15	0.88	0.98 ± 0.02	-2.67	0.92	16.00	5.79 ± 2.91
Founders	1156	379	0.33	0.94 ± 0.05	-4.65	0.65	13.92	4.20 ± 2.09
Ancient Britain	48	36	0.75	098 ± 0.07	-3.52	0.90	16.18	5.54 ± 2.70

Table IV.6 Mitochondrial DNA diversity values for the ancient settlements and several groupings, divided according to the different periods. The *early Saxon* group includes the settlements of Norton, Buckland and Lavington (5^{th}–7^{th} century) and the *late Saxon* group includes only the Norwich settlement (9^{th}–11^{th} century). The *founder* sample included putative source populations (England, Wales, Cornwall, north Germany, Denmark and Norway). Statistics are as in previous table.

Individual archaeological sites also showed high levels of haplotype and nucleotide diversity when analysed separately (see Table IV.6). For the ratio k/N, all archaeological sites, except for Buckland, showed values even higher than the ancient samples taken as a whole. For the genetic diversity (GD) the Buckland and Lavington settlements showed values a little lower, which was also reflected in the Early Saxon group. For the standardized Shannon index, all the sites, except for Lavington, showed values much higher than the ancient samples as a whole. Same was found for the nucleotide diversity. In the case of the haplotypic richness, due to their too small sample size, archaeological sites could not be analysed separately. According to their historical period (using a g = 18), all the groups showed high values of haplotypic richness, which compared to the modern populations (see Table IV.5) would place them, as for the ancient sample as a whole, among the ten highest values.

On the other hand, the group of possible founder populations showed low haplotype diversity, with k/N = 0.33, GD = 0.94 ± 0.01, H'_S = 0.65, $r_{(18)}$ = 13.9 and $r_{(48)}$ = 32.55 All these estimators were comparatively lower (with GD significatively lower, $P = 6.7\ 10^{-5}$) than the ones obtained for both the ancient samples as a whole and for the individual archaeological sites. Interestingly, these values were very similar to those obtained for the modern English population (k/N = 0.45, GD = 0.93 ± 0.01 H'_S = 0.70 and $r_{(48)}$ = 31.44).

Differentiation between groups divided according to their historical period was calculated by AMOVA analyses (Excoffier *et al.*, 1992). Variation within groups accounted for >96% of the total genetic variance. F_{ST} between the different communities were not significantly different from zero (only Norwich was different, yet when the sequential Bonferroni correction was applied it remained non significant).

IV.2.4 Demographic analyses

IV.2.4.1 Theta estimators

Three θ estimators were calculated for the ancient and the modern populations. Each of these estimators uses different aspects of the genetic data to estimate $2N_{fe}\mu$, where N_{fe} represents the female effective-population size and μ the mutation rate. Because the mutation rate should be the same in all populations, differences in θ values should reflect differences in female effective population size.

θ_π estimates the effective female population size that would have allowed the observed number of pairwise differences between sequences to occur through mutation in a single population. This means that it tends to reflect N_{fe} over long periods of time and is therefore strongly affected by demographic fluctuation (Helgason *et al.*, 2000). Further analyses of this aspect will be shown later. θ_S is based on the relationship between sample size and the number of segregating sites, while θ_k is based on the relationship between the sample size and the number of observed haplotypes. These two measures are thought to be more sensitive to the effects of lineage sorting (i.e. the redistribution of lineages among different populations due to migration) during recent demographic history (Helgason *et al.*, 2000).

Estimators were calculated for the ancient samples and the modern population of England (see Table IV.7). The three θ exhibited considerable disparity among the modern populations and between the θ values themselves, varying up to one order of magnitude. θ_π showed no correlation with either θ_k or θ_S ($r = 0.18$ and $r = 0.10$) but θ_k and θ_S were strongly correlated ($r = 0.91^{**}$, $P = 0.01$). Similar relationship between the θ estimators have been previously reported (Helgason *et al.*, 2000, 2001).

Population	N	$\theta_\pi \pm$ sd	$\theta_S \pm$ sd	θ_k (95% CI)
Leicester	6	6.08 ± 3.89	6.13 ± 3.21	n/a
Norton	12	5.69 ± 3.29	5.62 ± 2.49	58.44 [14.4 - 252.3]
Buckland	7	5.27 ± 3.31	4.89 ± 2.52	6.39 [1.7 - 24.4]
Lavington	6	2.37 ± 1.73	2.19 ± 1.34	11.44 [2.6 - 52.0]
Norwich	17	5.79 ± 3.26	7.09 ± 2.82	57.22 [18.4 - 196.2]
Period				
Romano-British	6	6.08 ± 3.89	6.13 ± 3.21	n/a
Early Saxon	25	4.88 ± 2.74	5.82 ± 2.19	44.25 [19.0 - 109.5]
Late Saxon	17	5.79 ± 3.26	7.09 ± 2.82	57.22 [18.4 - 196.2]
Ancient Britain	48	5.50 ± 2.99	8.33 ± 2.65	63.76 [34.8 - 120.3]
Modern England	344	3.81 ± 2.12	13.40 ± 2.98	106.5 [85.1 - 133.1]

Table IV.7 Three different estimators of the female effective population size (N_{fe}), based on the number of mutational differences among the sequences (θ_π), on the number of segregating sites (θ_S) and on the number of haplotypes (θ_k).

The θ_k value (i.e. the effective population size as derived from the number of haplotypes observed) obtained for modern England was approximately two times the value for ancient England (~106 and ~64, respectively). However, θ_π (i.e the effective population size as derived from the nucleotide differences among the sequences) was significantly higher ($P = 0.00003$) for the ancient population. The three estimators were also calculated for each archaeological site separately. The small sample sizes did not allow many interpretations. The effective female population size estimated for the Norton cemetery (θ_k ~58) is remarkably close to the estimated number of females based on archaeological data: the total number of individuals excavated at the Norton cemetery was ~120 and it is believed that both sexes were equal proportions (Marlow, 1992). It should be noted however that this is not the female population size of the settlement of Norton. At the Norwich cemetery several skeletal remains were fragmented, so that osteological analyses allowed sexing only 65 of the total 107 remains. These showed a higher proportion of women (41), yet these numbers cannot be

extrapolated to the entire population (Popescu, in prep.). Still the θ_k value obtained seems to be in the expected range. For the other three populations the sample sizes were too small to make any inference.

IV.2.4.2 Mismatch distributions

Pairwise comparisons of differences between individuals were calculated using the software package DnaSP 3.53. The program retrieves the distribution of the observed pairwise nucleotide site differences (called a mismatch distribution) and the expected values for stable vs. expanding population models. The modern and ancient English samples were tested against these two models. Goodness of fit to both curves is shown in Figure IV.7. Neither of the populations fit the curve for a constant size population but instead the curve for expanding populations.

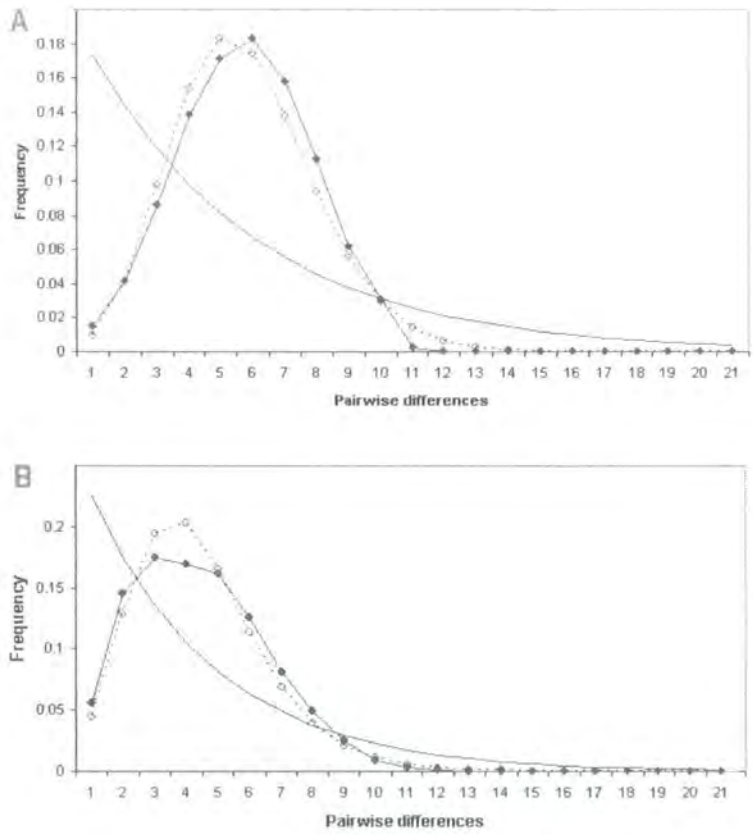


Figure IV.7 Mismatch distribution of pairwise differences for: **A)** ancient and **B)** modern English sample. Full line: distribution expected for constant size population; clear circles: distribution expected for populations under expansion; dark circles: observed distribution.

Tajima's *D* statistic (Tajima, 1989) was also calculated for these two populations. Although the test is originally designed to test for deviations from neutrality, it can also provide information about other demographic processes. A change in population size can result in deviations from neutral patterns of nucleotide variation of populations at equilibrium (for which *D* values approximate to zero). For example, negative values can be expected for populations under expansion, where due to the increase in population size there is a temporary excess of new mutations segregating at rare frequencies (Fay & Wu, 1999). Whereas the hypothesis of constant size cannot be fully rejected for the ancient sample (D = -1.6; NS, 0.1> P >0.05), the modern English sample showed significantly negative values (D = -2.2**, P <0.05).

In a case of population growth the resulting curve will travel at rate 2μ and its crest will be at $\tau = 2\mu t$ after t generations (Rogers & Harpending, 1992; Harpending *et al.*, 1993). Thus, the location of the wave provides a rough idea of the time (in units of $1/2\mu$) since the episode of population expansion. The mismatch distributions of several modern populations are shown in Figure IV.8. In general, they overlapped one to another, yet a pattern with respect to their relative position was observed. The crests of mismatch distribution of populations from the Near East, such as Palestine and Syria, and also the sample from the north European Russia population were to the right of the *x* axis. Their crests were at larger values of pairwise differences, i.e. more mutations had been accumulated over the time, indicating earlier expansions. On the other hand, the crests (τ) of populations from the Iberian Peninsula such as Portugal, as well as populations from the British Isles, such as Wales, modern England and Scotland plotted closer to the left side of the graph, indicating more recent expansions. Norway also showed up at this side of the chart. Other populations were somewhere in between these two positions. Interestingly, the mismatch distribution of the ancient English sample is to the extreme right of the graph, indicating an earlier expansion. The mismatch distributions of ancient and modern England are shown separately in Figure IV.9. The crest of the mismatch distribution of modern England (τ = 2.438) is shifted to the left of the crest of the ancient England mismatch distribution (τ = 4.775), theoretically indicating a more recent expansion for modern England.

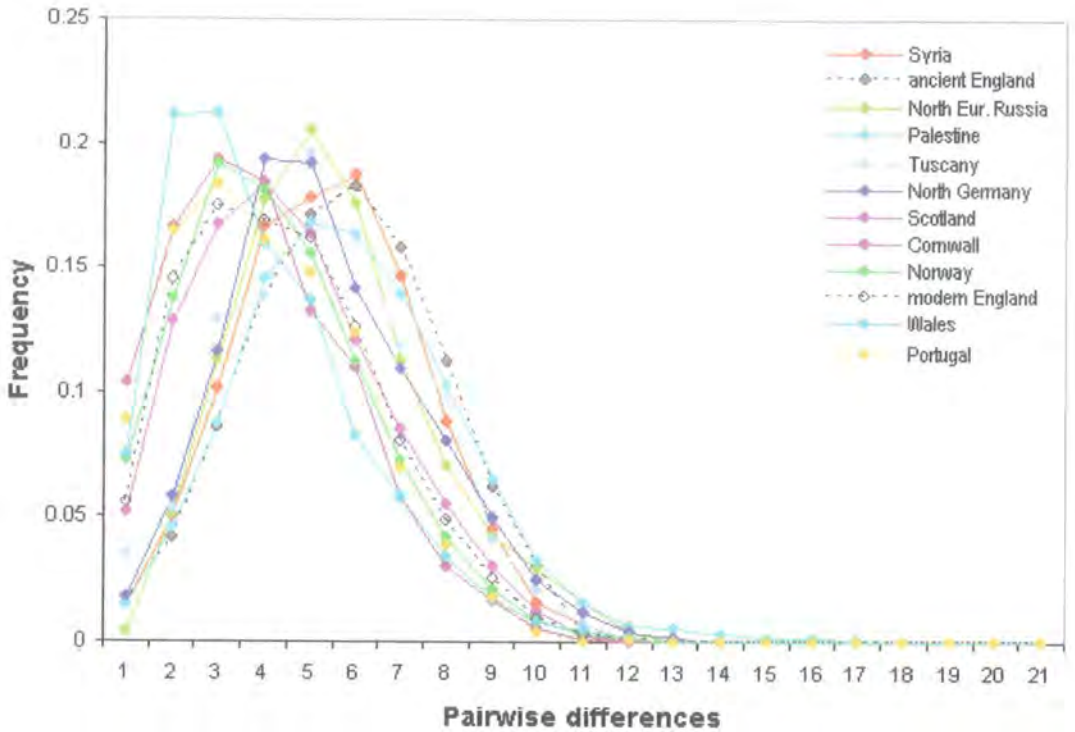


Figure IV.8 Mismatch distributions of pairwise differences for several modern populations. Codes for populations are shown in approximate order of plotting from right to left of the graph.

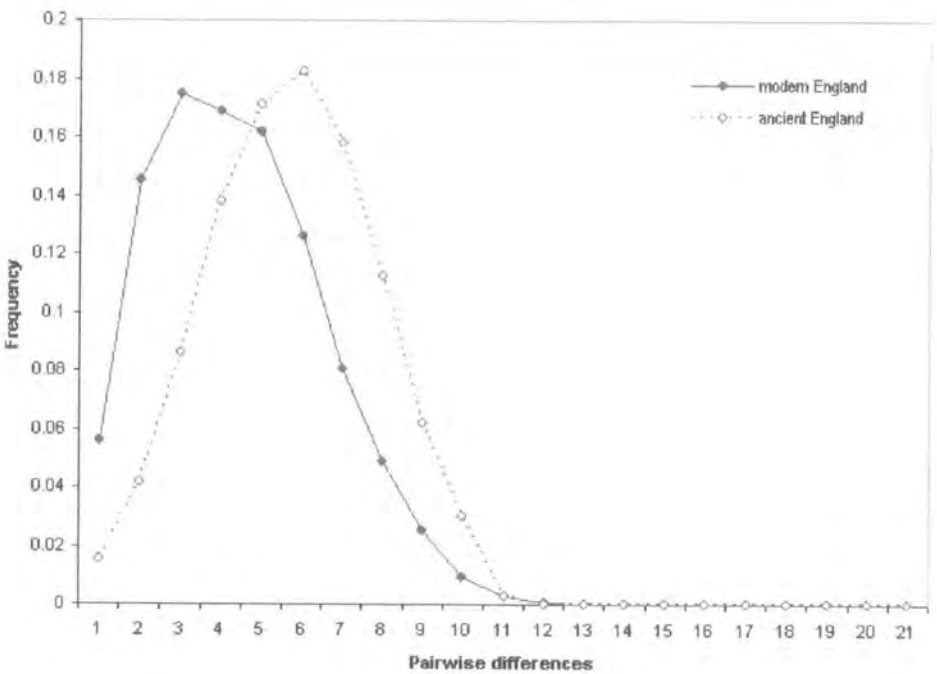


Figure IV.9 Mismatch distributions of pairwise differences for the ancient (clear circles and dashed line) and modern (dark circles and full line) English populations.

IV.2.4.3 Expansion times

The τ parameter was used to estimate times since expansion according to $\tau = 2ut$, where u is the total mutation rate for the whole fragment analysed. Several estimations of the mutation rate (μ) proposed in the literature, including both phylogenetically- and pedigree-derived approaches, have been used to estimate a range of expansion dates. Results for the ancient and modern English population as well as other comparative populations are shown in Table IV.8. As predicted by the location of the mismatch distribution curve, the ancient population showed expansion times much earlier than the modern English population. If average values are considered, the expansion of the ancient population is in the order of the expansion of the AMH to Europe. For the other populations, expansion times showed a trend which approximately decreased from East to West and to a lesser extent from North to South and has been observed previously for other populations (Comas *et al.*, 1997).

It should be noted that in addition to the large discrepancy in mutation rate estimates, the short highly variable HVS-I fragment analysed in this study results in an overestimation of the expansion times so that they should be seen as an indicator of the relative times.

			Expansion times (KY)			
			mutation rate (μ)			
Population	τ	r	$8.65\ 10^{-8}$ (a)	$1.03\ 10^{-7}$ (b)	$1.8\ 10^{-7}$ (c)	$4.5\ 10^{-7}$ (d)
Syria	4.471	0.016	125	105	60	24
Uralic Russia	4.229	0.026	118	99	56.5	22.5
north Italy	4.198	0.011	117	98.5	56	22.5
north Germany	3.913	0.016	109.5	92	52	21
Turkey	3.838	0.013	107	90	51	20.5
Is. Skye	3.046	0.016	85	71.5	41	16.5
modern England	2.438	0.014	68	57	32.5	13
Spain	2.312	0.015	64.5	54	31	12.5
Norway	2.241	0.013	62.5	52.5	30	12
Basque Country	2.123	0.023	60	50	28.5	11.5
Wales	1.761	0.026	49	41.5	24	9.5
ancient England	4.775	0.014	133.5	112	64	25

Table IV.8 Tau parameters and estimated expansion times, in thousand years (KY). Different mutation rates were used following: (a) Vigilant et al. (1991), (b) Horai et al. (1995) and (c) Forster et al. (1996) and (d) Howell et al. (2003). This latter is a pedigree-derived estimate. Mutation rates are per site per million years.

IV.2.5 Phylogenetic analyses

IV.2.5.1 Reliability and resolution

Phylogenetic analyses were performed using reduced median networks (RMN) (Bandelt *et al.*, 1995), in preference to traditional tree-construction methods, such as maximum parsimony (MP) or neighbour-joining (NJ). These methods which were originally developed for interspecific data are not appropriate for intraspecific data, in which the sequences are very closely related. In a RMN, genetic variants (mtDNA haplotypes in this case) are shown as inter-linked nodes of a network rather than ending tips of the branches of a tree. Thus, RMNs display the principal relationships present in the data and resolve likely parallel events while retaining character conflicts in the form of reticulations when ambiguity remains.

As discussed in the Introduction, Richards et al. (1996) were the first to use this approach to infer phylogenetic relationships of sequences from the HVS-I data from European populations. Among 821 HVS-I sequences analysed, 82 (10%) were observed more than once. Only these 82 haplotypes were used by the authors to construct their network. In their star-like phylogeny most of the sequences were close to the central node, yet some protruded as branches which shared a common ancestral segregating site. In this manner, the original six clusters or haplogroups (hg) were identified.

Before undertaking the phylogenetic analyses of the ancient and modern samples under study, some practical concerns were addressed, a) the reliability of the median construction method to be used, b) the possible loss of resolution due to the relatively short length of the HVS-I to be analysed, and c) the power of identifying the established haplogroups based on a different HVS-I data set.

The construction of relatively neat median networks is not always straightforward. To properly resolve the excess of reticulations it is usually necessary to let nucleotide positions showing a high mutation rate (hot spots) appear more than one time in the network. For this, the weighting scheme proposed by Richards et al. (1998) and Helgason et al. (2000) and based on (Hasegawa *et al.*, 1993) was implemented. Polymorphic sites were divided into three groups according to their mutation rate: fast, average and slow. In this manner, nucleotide sites with high mutation rates (which are

therefore likely to have mutated more often) will be present more than once in the network allowing a better resolution of the reticulations. If the hypercube structure still remained, the reduced median (RM) and neighbour joining (NJ) method were applied sequentially in order to further clarify the network, as suggested by the software manual. Thus, to test whether this method would yield accurate and reliable networks, the raw data from Richards et al. (1996) was used to reproduce their results. The network obtained is shown in Figure IV.10 as well as the original network for comparisons. The two networks showed a similar topology with the same identifiable groups. Thus, the method used in this study allowed the reproduction of previous studies.

Obtaining a median network of well-resolved clusters clearly depends on the number of segregating sites, and therefore the length of the sequence analysed and used for its construction. Richards et al. (1996) identified the six major haplogroups, based on a ~270 bp fragment of the HVS-I (although Sykes (1999) referred to it as a 350 bp fragment; the origin of the discrepancy is not known). To test whether resolution would be lost in the analyses, the raw data by Richards et al. (1996) was trimmed to the size analysed in this study (207 bp). The reduced median network (RMN) obtained with this shorter fragment is shown in Figure IV.11. Although this network is based on fewer segregating sites, it still allowed the resolution of the majority of the original clusters, except that the absence of segregating site 069 resulted in a poorer internal definition of group 2. It should be noted, however, that the original data did not include site 069 for all sequences either, yet the authors included it in their RMN.

In comparison with more recent phylogenies including more HVS-I and also HVS-II data (see Introduction, Fig. I.12), the phylogeny shown in Figure IV.11 lacks segregating sites 069, 343, 356 and 391 from HVS-I and 073 from HVS-II (which are beyond the length of the fragment analysed in this study) and therefore some of the clusters cannot be identified. Haplogroup J (former 2A) can be defined as super-hg JT (i.e. presence of 126) minus hg T (i.e. presence of 294) and hg I can still be defined by the presence of 223 and 129. However, hg U3 and hg U4 (defined by sites 343 and 356, respectively) cannot be differentiated. This must be taken into consideration during the phylogenetic analyses of the samples.

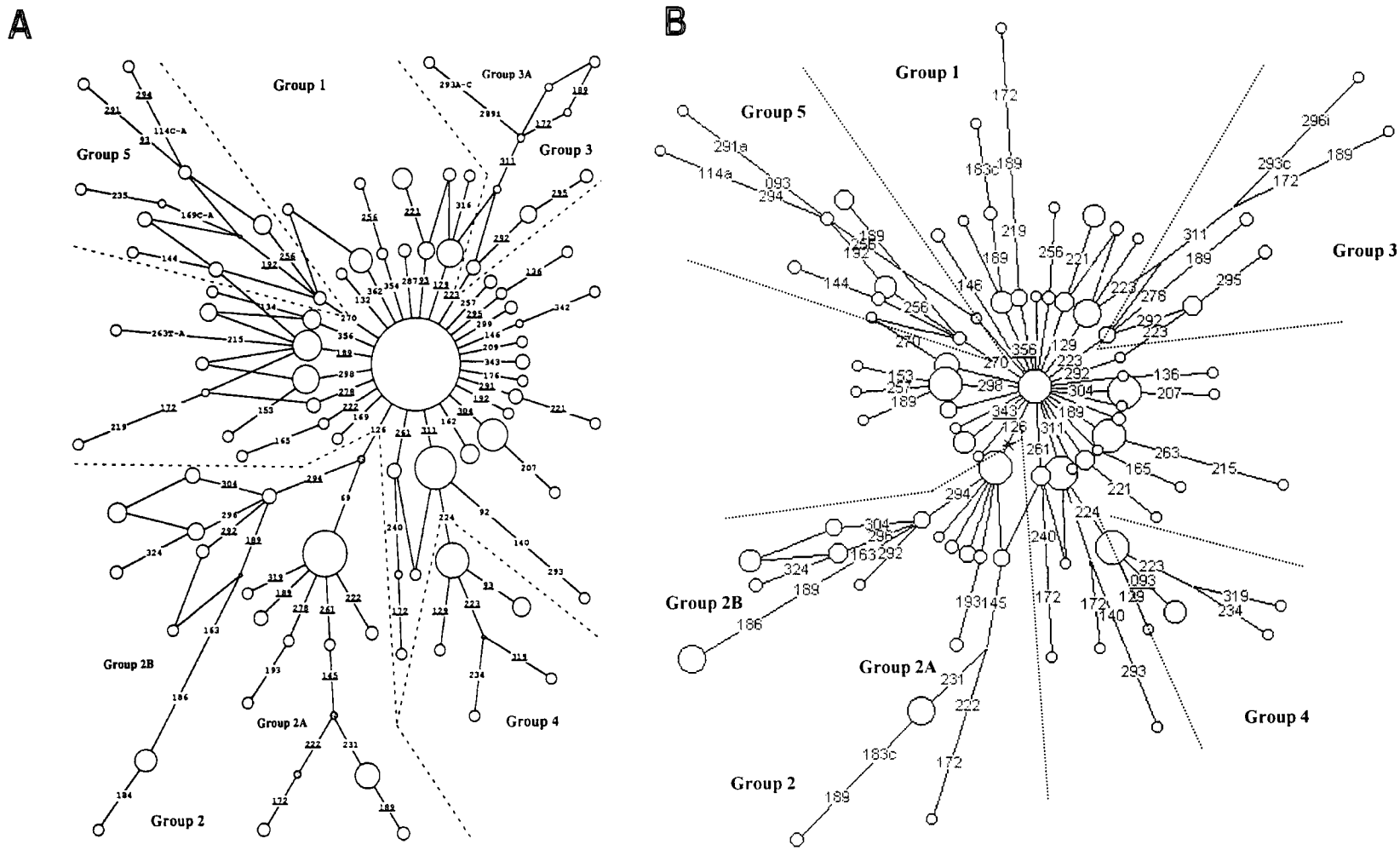


Figure IV.10 Reproducibility of reduced median networks (RMN) for European data. (**A**) Original RMN as published in Richards *et al.*, (1996) and (**B**) Reproduced RMN based on the raw data from Richards *et al.* (1996), using NETWORK 2.0 (Rölh, 2001) and the weighting scheme based on Hasegawa *et al.* (1993). Segregating positions are indicated. The major haplogroups are shown in both diagrams (the old nomenclature was used, following the original publication). NOTE– Site 069 (indicated with a cross) near the root of group 2 in diagram (**B**) was not reported for all the original data although the authors included it in their RMN.

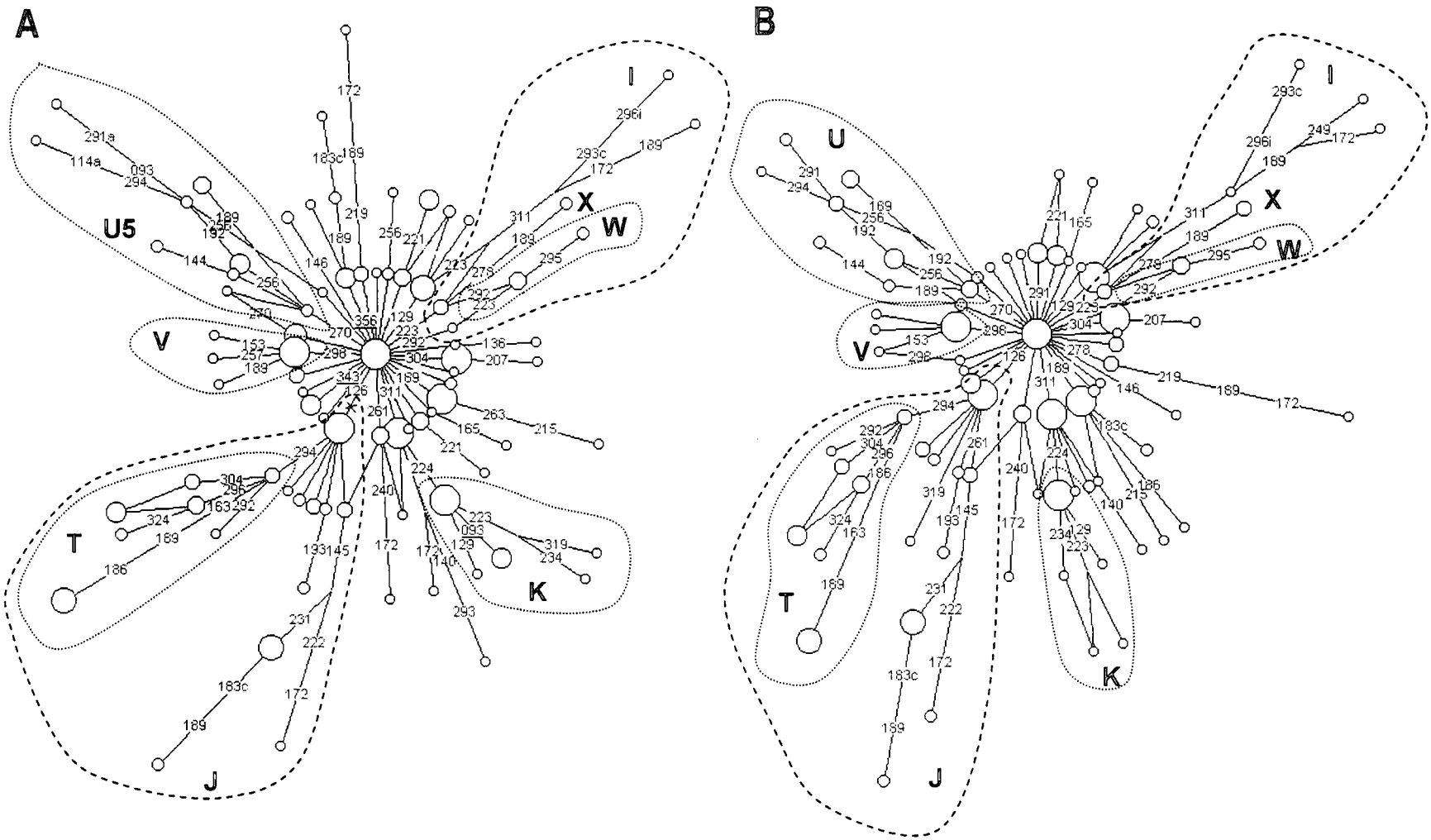


Figure IV.11 Comparison of resolution for networks based on different sizes of HVS-I regions. (**A**) Reproduced reduced median network of 270-bp-sequences from Richards *et al.* (1996), and (**B**) Reduced median network obtained for the same data from Richards *et al.* (1996) but trimmed to 207 bp. Segregating sites are indicated. Major haplogroups are shown in both diagrams (the current nomenclature was used). NOTE– Position of site 069 is indicated as before.

Lastly, analyses were carried out to determine if the same original or also new haplogroups could be identified in a network constructed based on my mtDNA database. Among the 6454 HVS-I sequences compiled, 1412 different HVS-I haplotypes were observed, of which 468 were multiple haplotypes (i.e. observed more than once). The total 1412 haplotypes were practically impossible to be run in the software used (Röhl, 1997) and the 468 multiple sequences were still far too many to produce a clear median network. Therefore, a different strategy was undertaken. The analysis was focused on the haplotypes best-represented in Europe since, on the assumption of neutrality, in a young expanding population, the more frequent the haplotypes tend to be older (Donnelly & Tavaré, 1986), and therefore generate a clearer topology. Thus, a network was constructed using haplotypes found at frequencies >10, comprising a total of 84 haplotypes (see Fig. IV.12). The original haplogroups could be clearly identified. In addition, other small protruding branches were observed, such as those defined by segregating sites 189, 304 and 162. These groups have previously been described as sub-haplogroups in other European networks (named sub-hgs H1, H3 and H8, respectively; Helgason *et al.*, 2001).

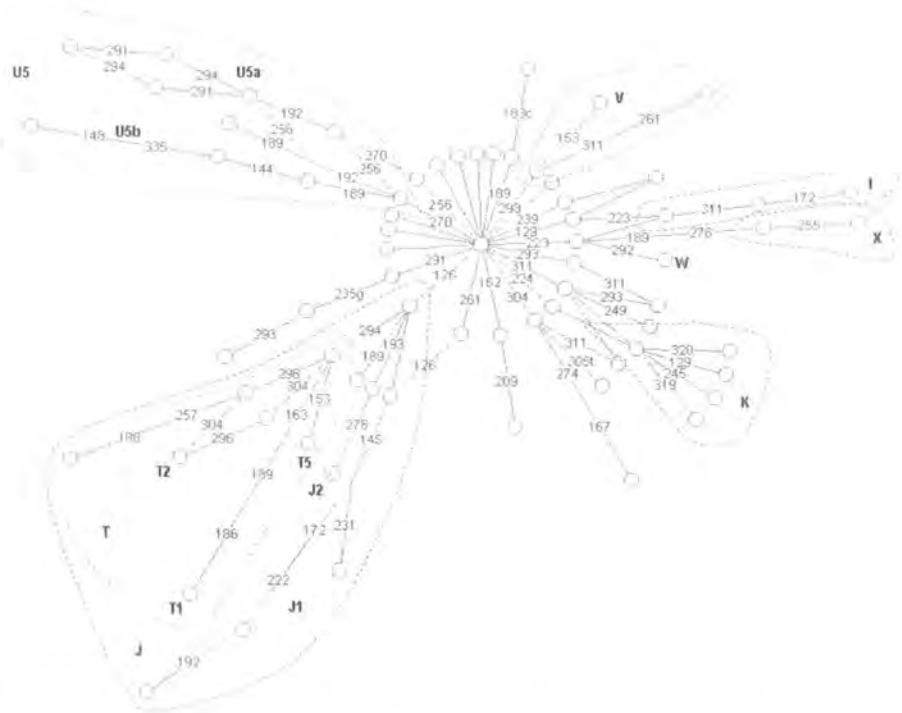


Figure IV.12 Reduced median network for European HVS-I sequences from the compiled database showing frequencies >10 (a total of 84 haplotypes). Haplogroups are indicated.

IV.2.5.2 Phylogenetic analyses of ancient Britain

The 48 HVS-I sequences obtained from the ancient samples were used to construct the reduced median network shown in Figure IV.13. The RMN showed a typical star-like phylogeny in which the CRS was the most common haplotype and the central node of the network, with branches protruding from it. This structure has been observed in the majority of networks drawn for individual modern European populations and is characteristic of expanding populations. Some of the haplotypes were within one or two mutational events of the central node, however, other haplotypes formed branches that protruded significantly beyond this limit, being up to 6 mutations apart from the CRS. All these branches corresponded with the groups observed by Richards et al. (1996) and also with the most detailed classification more recently adopted. Hence, despite the relatively restricted resolution, all the major original European haplogroups, that is I, J, K, T, U5, V, W and X, were identified in the network. Some of their sub-haplogroups, such as T1, T2, T4 and T5 and U5a (and also U5a1) and U5b were also identified.

Most of the ancient HVS-I haplotypes could be positively assigned to one of these major clusters or their sub-clusters as they presented the corresponding key segregating sites. Some of the ancient HVS-I haplotypes were private to the ancient population of Britain, i.e. they were not found in any other population among the compiled mtDNA database (see Chapter V). These are further indicated in Figure IV.13.

Excluding hg H, which will be discussed later, hg T was the best-represented haplogroup among the ancient samples (~23%), including haplotypes belonging to its sub-hg T1, T2, T4 and T5. Haplogroup U5, defined by segregating site 270, including both sub-hg U5a and U5b, was the second most common hg (~17%) among the ancient samples. Four sequences (~8%) belonged to super-haplogroup JT, defined –within the analysed HVS-I fragment analysed in this study– only by 126, but none to its sub-group J1. Two haplotypes belonged to hg V (~4%) and one to hg K (2%). Haplogroup I included one haplotype (2%) and hg W and hg X both two haplotypes (~4%). Regarding this latter, sequence w6 presented two of the segregating sites which define hg X, position 223 and 278, but at position 189 instead of the characteristic T→C, haplotype w6 presented a rather unusual transversion T→A. It might be claimed that this haplotype therefore should not be classified as hg X. However, other states of site 189

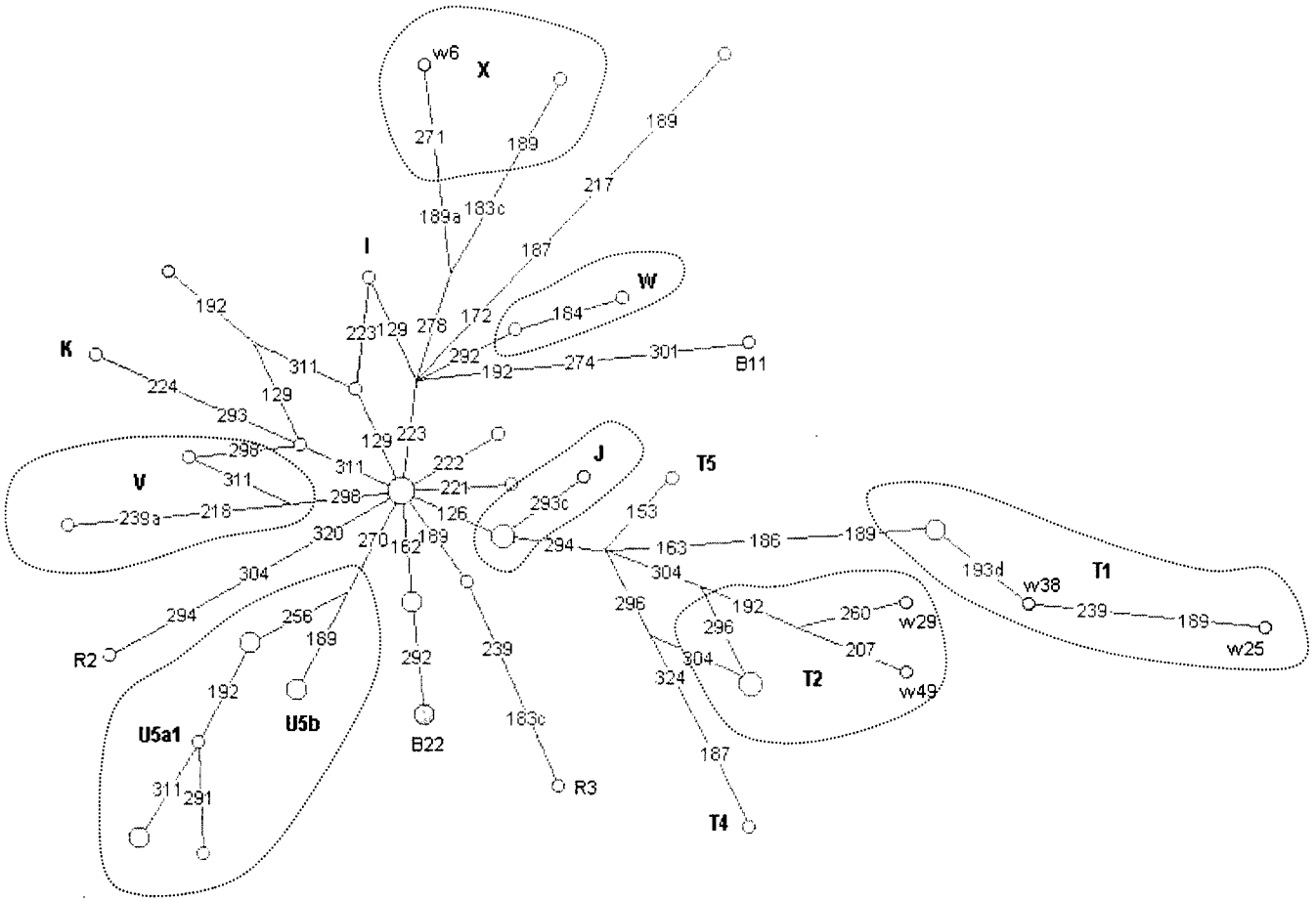


Figure IV.12 Reduced median network for the 48 HVS-I sequences from the ancient population of Britain. Haplotypes are represented by nodes and their size is proportional to their frequency. Dark nodes represent haplotypes that are private to the ancient population of Britain. Segregating sites are defined as differences with respect to the CRS (Anderson *et al.*, 1981). All sites are transitions unless further indicated (e.g. 239a: adenine transversion and 193d: deletion). The major haplogroups and sub-haplogroups are indicated. Individual samples mentioned in the text are further indicated.

such as the wild-type nucleotide T (i.e. absence of the segregating site at all) or this particular transversion, have been previously observed in haplotypes which were positively classified as hg X by RFLPs analyses by Richards et al. (2000) and Gresham et al. (2001). The two sequences w5 and B11 were close to haplogroups I, X and W and share segregating site 223, but could not be unambiguously classified into any of the European haplogroups. Sample w5 shared some of its remaining segregating sites (172, 187 and 189) with sequences that have been classified as haplogroup B on the basis of RFLP data (presence of a 9bp deletion) (Richards *et al.*, 2000). Such classification for w5 would be only conjectural. Sequences similar to B11 have not yet been observed in the available HVS-I database. In any case, both w5 and B11 did cluster into the super-group defined by the presence of the ancestral C→T transition at position 223, which was also very well-represented overall (~14%).

The remaining 11 haplotypes did not present any defining segregating site for the major haplogroups and therefore could not be positively classified into any of those clusters. Most of them were only one or two steps divergent from the CRS and might belong to hg H. In particular, haplotypes B22 (which is the same as B31), R3 and R2 might belong to the above mentioned small sub-hg H8, H3 and H1, respectively. It must however be remembered that, since site 073 of HVS-II was not included in the analyses, classification as hg H cannot be at all conclusive as haplotypes from hg U3 or hg U4 might yield similar sequences.

Figure IV.14 shows the same median network where the archaeological origin and period of the samples is indicated. Samples from all archaeological sites were widely dispersed across the network. However, it is worth noticing the prevalence of Norwich samples in hg T. Two of the branches, including T1 and T2 haplotypes, were formed almost exclusively by six Norwich samples. These samples differed one to another by one or two sequential mutations. In addition, four of these haplotypes were unique in the 6454 HVS-I database. On the other hand, sub-hg U5a was mainly represented by Norton and Lavington samples. Moreover, the only two hg W haplotypes were also exclusive of the Norton settlement.

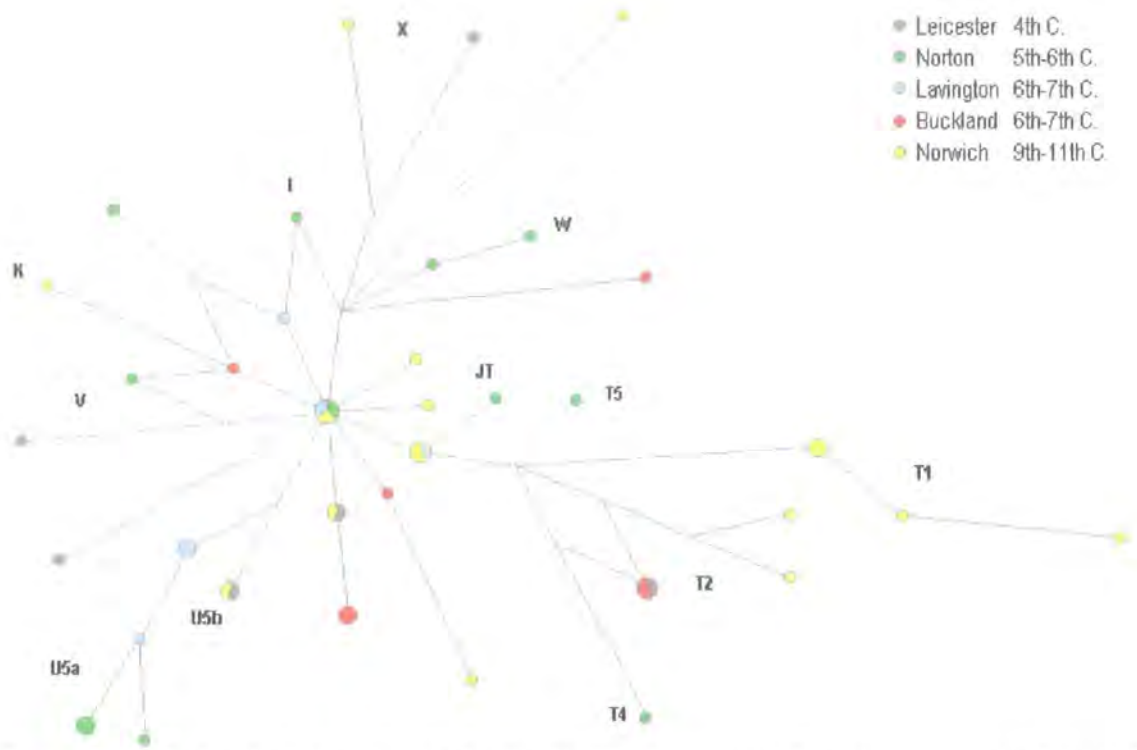


Figure IV.14 Reduced median network for the 48 HVS-I sequences from the ancient population of Britain, showing their respective archaeological settlement and period as indicated in the key. Nodes are proportional to the number of individuals. Major haplogroups are indicated as before.

IV.2.5.3 Phylogenetic analyses of modern England

To construct the RMN for the population of modern England 258 sequences were used (sequences from Cornwall were not included). Figure IV.15 shows the full reduced median network for all the 258 modern English sequences. It showed, as for most European networks a star-like structure, with the CRS being by far the most common sequence and the central node (given the resolution of the analysed segment). As for the RMN for the ancient population, all the major European haplogroups and sub-haplogroups were observed. However, their relative representation was different. Haplotypes related to the CRS and possibly pertaining to hg H accounted for >50% of the haplotypes. The next well-represented haplogroup was hg T (determined by segregating sites 126 and 294), yet with only a ~9% of the total modern English samples, as opposed to the ancient sample with 23%. Haplogroup U5 was also present at a much lower frequency than in the ancient sample (~8% vs. 17%), as were hg X (1.2% vs. 4.2%) and hg W (1.2% vs. 4.2%).

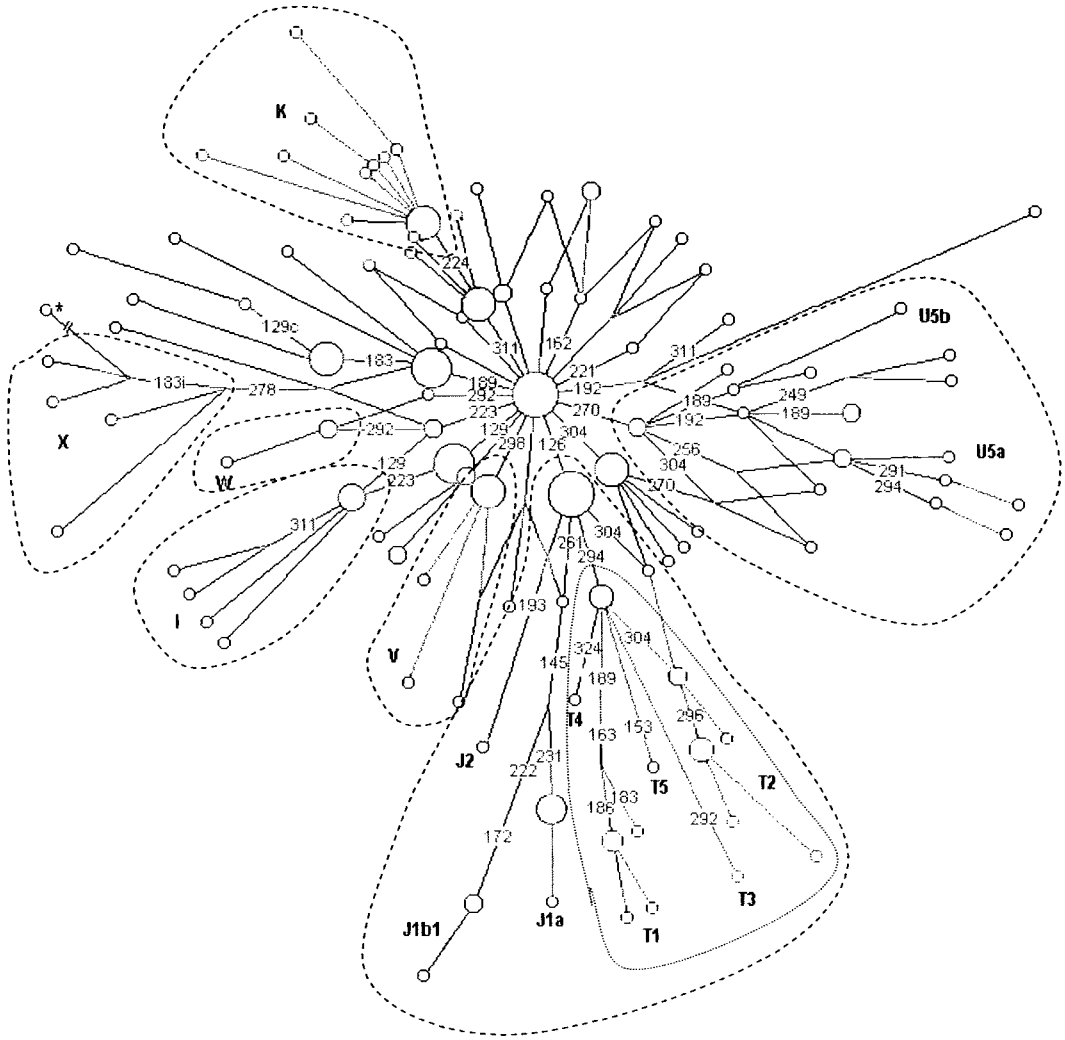


Figure IV.15 Reduced median network for the 258 HVS-I sequences from the modern population of England. Haplotypes are represented by nodes and their size is proportional to their frequency. Length of branches connecting nodes is proportional to the number of mutations (although only relevant sites are indicated). Segregating sites are defined as differences with respect to the CRS (Anderson *et al.*, 1981). Major European haplogroups and sub-haplogroups are shown. NOTE– (*) indicates an African sequence which is 5 mutations away from the closest node.

On the other hand, hg K was more frequent in the in the modern English population (6.5% vs. 2%). Haplogroup V was approximately equally represented in ancient and modern populations (4.2% and 3.9%, respectively), as was hg I (2.1% and 2.3%, respectively). Few (0.8%) related HVS-I sequences clustered under the ancestral transition C→T at 223 which accounted in total only for less that 6%. Overall, the distribution of haplogroups in the ancient and modern population of England was significantly different ($\chi^2 = 42.77$, d.f. = 26, $P = 0.02$)

IV.2.5.4 Phylogenetic analyses of individual haplogroups

In order to further analyse the phylogenetic relationship between the ancient and modern samples of England, HVS-I sequences from both populations were analysed together. They comprised a total of 306 HVS-I sequences (48 and 258, respectively), representing 150 HVS-I haplotypes. Of these, 14 haplotypes (9.3%) were shared between both populations. The majority of the shared haplotypes were haplotypes common among European populations, such as CRS, [126], [129], [162], [189], [221], [311], [126-163-186-189-294] (defining sub-hg T1), [126-194-196-304] (defining sub-hg T2), [189-270] (defining sub-hgU5b), [192-256-270] (defining sub-hg U5a1), [223-292] (defining hg W) and haplotype [192-256-270-291] an haplotypes mostly found among northern Europeans (Richards *et al.*, 1996; 1998; Macaulay *et al.*, 1999). However, haplotype [294-304-320] –observed in an ancient sample belonging to the Roman cemetery at Leicester– was rare: shared between ancient and modern England and Germany (see Chapter V and Appendix F).

The RMN constructed using the >300 HVS-I sequences presented too many reticulations. For clarity, RMNs were then constructed for the different haplogroups, which grouped the HVS-I sequences sharing the respective segregating sites. RMNs were constructed for the hg or super-hg for which there was a significant number of ancient HVS-I sequences, namely JT, U5 and IXW.

Super-haplogroup JT

Figure IV.16 shows the RMN obtained for super-hg JT for the ancient and modern population of England, comprising 22% of the total haplotypes (and 23% of the sequences). By definition, the central node of the network was the ancestral haplotype 126. The RMN showed a star-like topology, with several long branches protruding from the central node. At one side of the core sequence, segregating site 294 defines hg T, representing ~64% of the haplotypes of super-hg JT. The other side (all the non-294 haplotypes) represent hg J. As observed earlier, only one hg J ancient sequence was present, whereas the modern samples showed a large proportion of J1 and in particular J1b1. This haplogroup is supposed to be British-specific (see Chapter V). On the other hand a relative large number of ancient haplotypes were hg T.

Only three haplotypes were shared between the modern and ancient samples, namely the ancestral haplotype 126 and haplotypes [126-294-296-304] and [126-163-186-189-294] both common haplotypes belonging to T2 and T1, respectively (see Chapter V). Two pairs of ancient haplotypes were outliers in the network, forming either a small group (in sub-hg T2) or a branch (in sub-hg T1) (see Fig. IV.13). They were connected to modern haplotypes by one mutational event. Another ancient sequence was the end (connected by one mutational event) of a branch which included mainly modern sequences. Only one sequence, haplotype [126-153-294], belonging to sub-hg T5, was an internal node connecting three modern English haplotypes.

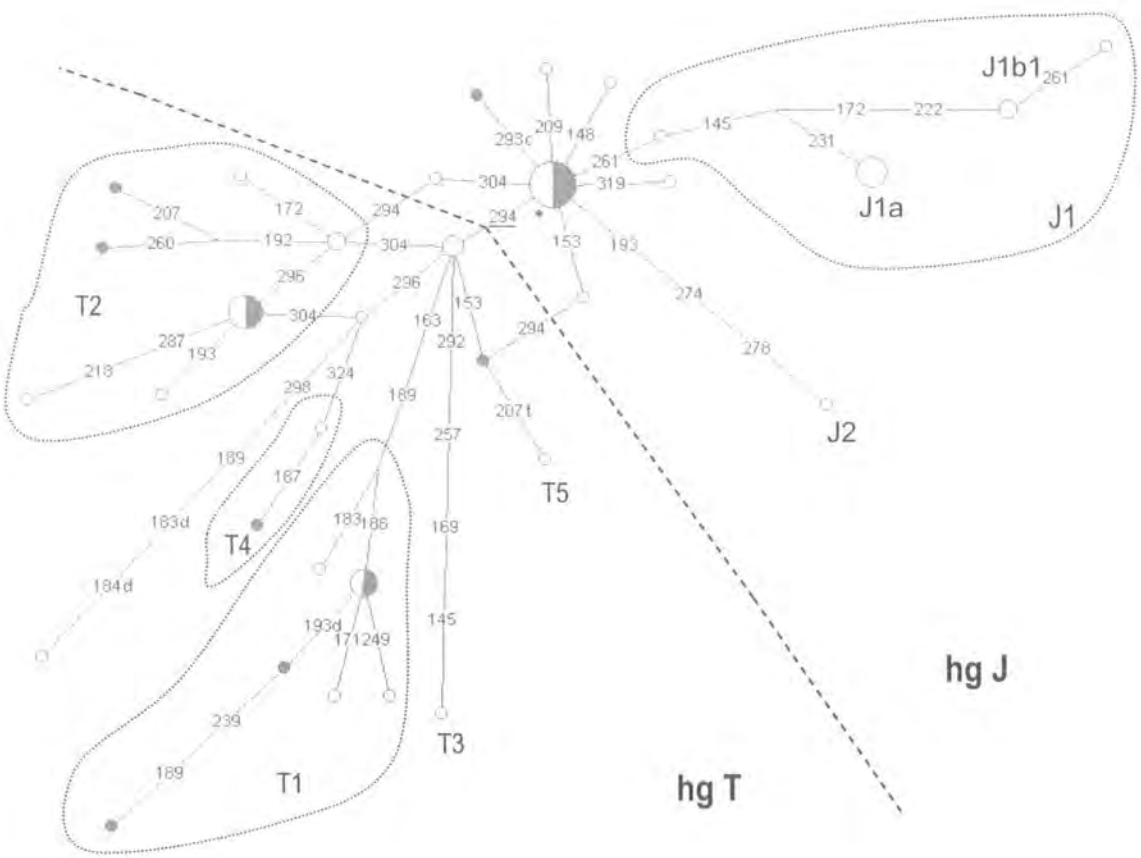


Figure IV.16 Reduced median network of super-hg JT, with central haplotype [126] indicated with (*). Segregating site 294, defining hg T is underlined. Sub-haplogroups are shown. Dark and clear nodes represent haplotypes from the ancient and modern population of England, respectively. Two-colour nodes are shared haplotypes.

Haplogroup U5

Figure IV.17 shows the RMN obtained for hg U5 for the ancient and modern populations from England. It comprises 19 haplotypes, accounting for 12.5% of the data (and 29 sequences, accounting for 9.5% of the sequences). By definition, in this case, the central node of the network was the ancestral haplotype 270. The topology of U5 was not star-like, with shorter branches and more ambiguities next to the central node. Nevertheless, the structure observed agreed with previously observed for the U5 cluster in Europe (Richards *et al.*, 1998), showing sub-haplogroups U5a (defined by site 192), U5a1 (defined by 192 and 256), U5a1a (an additional sub-group of U5a1a which lacks transition 192, and is therefore defined only by segregating site 256) and U5b (defined by site 189).

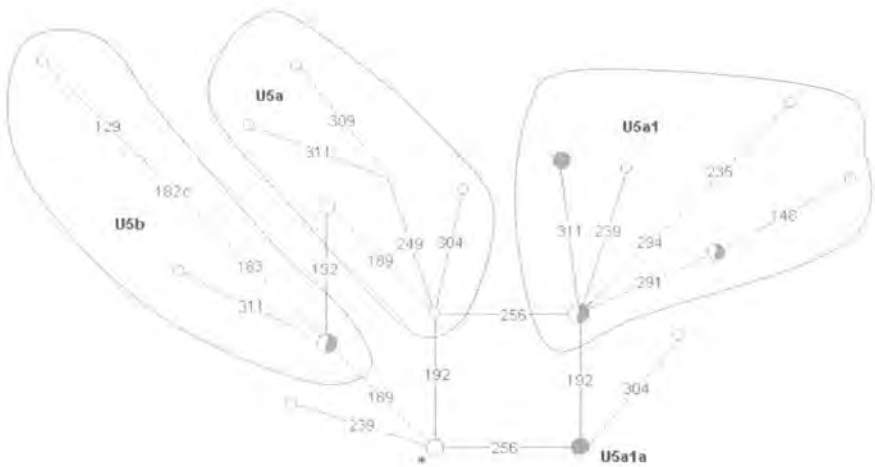


Figure IV.17 Reduced median network of hg U5, with central haplotype [270] indicated with (*). Sub-haplogroups are shown. Dark and clear nodes represent haplotypes from the ancient and modern population of England, respectively. Two-colour nodes are shared haplotypes.

Three ancient HVS-I hg-U5 haplotypes (60%) were shared with the modern HVS-I population. These are the main haplotypes defining U5a1 [192-256-270] and U5b [189-270] and another sequence type of U5a1 [192-256-270-291]. Also within U5a1 two ancient sequences were observed that were not shared with the modern population. One of them [256-270] defined the small sub-group U5a1a, which has been only found in northern Europeans. In general, ancient hg U5 haplotypes were more internal nodes of the network, as opposed to the case of the previous network in which they were at very ends of branches.

Super-haplogroup IWX

Figure IV.18 shows the RMN obtained for super-hg IWX for the ancient and modern populations from England, which comprises 12.5% of the haplotype and 7.5% of the haplotype and sequence data, respectively. By definition, in this case, the central node of the network was the ancestral haplotype defined by the transition C→T at position 223.

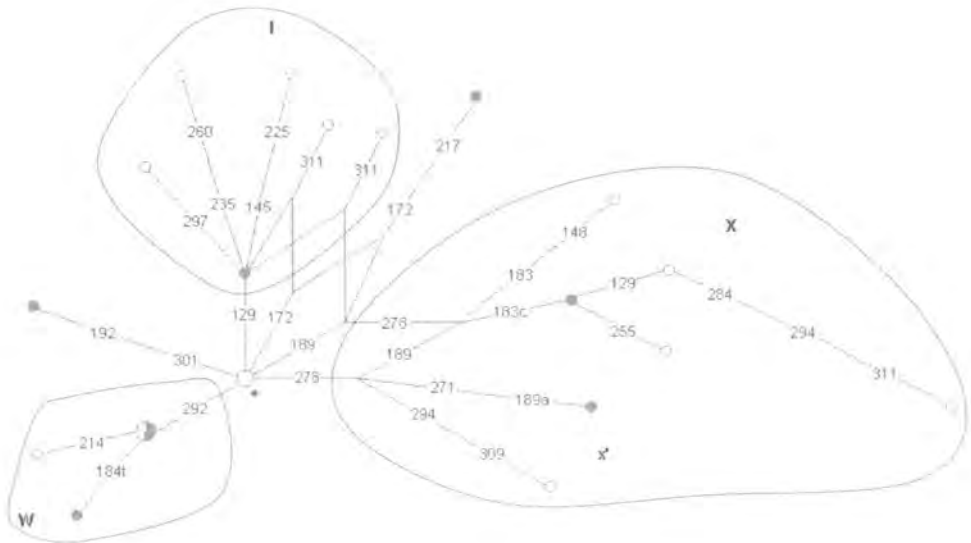


Figure IV.18 Reduced median network of cluster IXW, with ancestral haplotype [223] indicated with (*). Haplogroups are shown. Dark and clear nodes represent haplotypes from the ancient and modern population of England, respectively. Two-colour nodes are shared haplotypes.

The three major clusters I, W and X could be clearly identified in the network, with hg I showing some reticulations and ambiguities next the central node. Two ancient HVS-I haplotypes present in the network could not be assigned to any of these major haplogroups. Regarding the haplotype sharing only one haplotype (~14.3%) was shared between the ancient and modern population. This was the main haplotype defining hg W [223-292]. Other two W-haplotypes were observed, one from the ancient and one from the modern English population. In hg I, the main haplotype (which for the fragment analysed in this study is defined by sites 129 and 223) was observed in the ancient population but not in the modern one. However, several modern hg I haplotypes protruded from that node. In the case of hg X, neither the modern nor the ancient sample presented the main haplotype of the haplogroup [189-223-278]. However, the earliest

and simplest hg X haplotype [183-189-223-278] observed in the haplogroup, belonged to the ancient population and was also the ancestral node for 3 modern sequences. Also among the hg X sequences, two haplotypes, one modern and one ancient (indicated by x' in the network) did not present the segregating site 189 as such; the modern sequence presented the wild type variant whereas the ancient sequence, as described previously, presented a transversion T→A. In general, as for the case of hg U5, most of the ancient IWX HVS-I sequences were internal nodes of the three major clusters (I, X and W) of this network.

IV.3 Discussion

In this Chapter the genetic diversity of communities from Anglo-Saxon England was studied. This was done in the context of modern populations, in particular modern England. Several approaches were undertaken to depict the genetic variation observed in these 1,700 to 1,000 year old sequences. For these results to be of value, authentication of the ancient genetic material was of crucial importance.

IV.3.1 Authentication of ancient DNA

Several criteria have been established to determine the authenticity of ancient DNA sequences (Cooper, 1997 ; Cooper & Poinar, 2000; Hofreiter *et al.*, 2001b). They include: biochemical assays to test for preservation, extraction and PCR controls, ancient DNA behaviour, quantification of number of templates, amplification from a second extract, reproduction by a second laboratory and a phylogenetic sense of the data obtained. Each criterion has been addressed in this work as follows

Biochemical assays to test for macromolecular preservation — *assays* of this type, such as racemization of amino-acid sequences (Poinar, 1976; Bada *et al.*, 1994; Hofreiter *et al.*, 2001b) are used to determine whether the state of preservation of a specimen is compatible with aDNA preservation. These analyses were not undertaken, as they were outside the aims and possibilities of this study. However, they are usually recommended as a preliminary analysis in order to avoid the destruction of highly valuable specimens (e.g. the Neanderthal sample), which was not required for my study material.

Extraction controls and PCR controls for contamination — DNA aerosols are the usual source of contamination in DNA analyses and especially ancient DNA work. To identify potential contaminating sequences, DNA sequences of lab members were obtained for reference. It was thereby identified that the usual source of contamination was lab-borne with DNA sequences of lab members, in particular my own DNA. However, contamination of unknown sources was also observed. Although extensive autoclaving and sterilization was carried out, small fragments of DNA might not have been totally removed from plastic-ware (Schmidt *et al.*, 1995). Other possibility is that samples were contaminated with skin cells of the excavators who handled the skeletal material, yet this is unlikely, as teeth were extensively cleaned, UV-irradiatiated and the outer parts discarded.

In any case, sequencing of all contaminated samples allowed the identification of most of the contaminating sequences. Contamination was detected in ancient samples even when the respective negative controls were clean. This result has also been previously reported in other ancient DNA studies and is thought to be due to a carry-over or enhancing effect of the aDNA extracts (Handt *et al.*, 1994; Kolman & Tuross, 2000). A very low number of contaminating molecules in the extraction or amplification reagents may not yield any amplification products as they may be absorbed to the plastic-ware and therefore cannot serve as template for the PCR. However, when the reagents are added to the aDNA extract, which may itself contain DNA from micro-organisms or other molecules such as sugars, these act as carriers by displacing the contaminating molecules from plastic surfaces and thus allowing the contaminating molecules to become available for amplification.

Ancient DNA behaviour — *a)* inverse correlation between size of amplification product and amplification efficiency, and *b)* indications of damage in the DNA. Optimisation of the PCR conditions allowed amplification of fragments of ~250 bp but PCR efficiency of larger fragments (350 to 400 bp) of either HVS-I or HVS-II was considerably reduced and amplification of whole mtDNA CR was impossible (see Chapter III). When larger fragments were actually amplified they always turned out to be a contamination.

Several ancient DNA extracts yielded artifact or chimeric sequences. This was observed as different extracts yielding *similar* yet not identical sequences, such as the

DNA extracts of samples w1, w18, and L11. In all these cases, a number of extra variable sites were observed in one of the extracts but not in other extracts of the same sample. Moreover, these chimeric substitutions were almost always C to T transitions. The error rate of the DNA *Taq* Gold polymerase is in the order of 10^{-5} (per nucleotide per cycle) (Lundberg *et al.*, 1991) (Applied Biosystem). Therefore, regular polymerisation errors cannot account for all the observed nucleotide differences. However, miscoding lesions in the initial DNA template induce the misincorporations of bases during amplification (Hansen *et al.*, 2001). It is known that aDNA undergoes catalytic processes resulting in apurinic sites and breakage of the DNA strands. In particular hydrolytic deamination of cytosine and its homolog 5-methyl cytosine to uracil and thymine, generating CG→TA transitions (see Fig. IV.19) has been specifically described for ancient DNA amplification (Hofreiter *et al.*, 2001a).

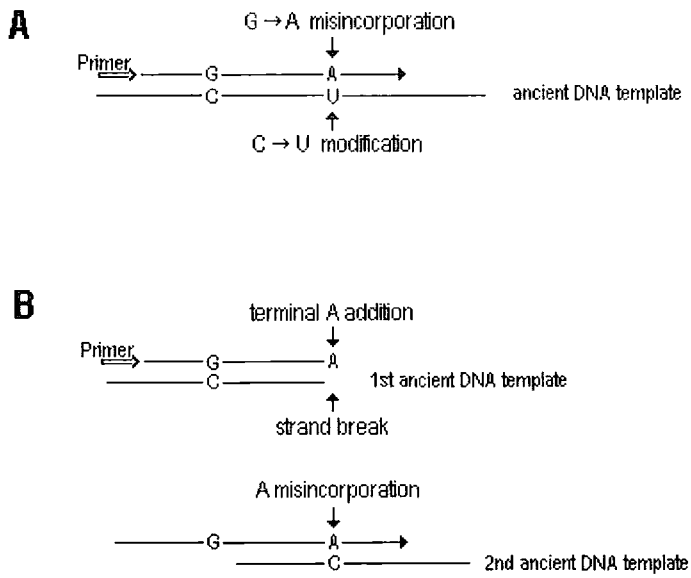


Figure IV.19 Proposed mechanisms explaining the incorporation of deoxyadenosine residues opposite to deoxycytidine in the template molecule. (**A**) Cytosine deamination. The U in the ancient DNA molecule template represents a deoxyresidue, a 5-hydroxydeoxyuridine residue or any other modified deoxycytidine residue read as T by the *Taq* polymerase. (**B**) Jumping PCR. In the first step, the primer is extended on a template that does not include the second primer site and a terminal deoxyadenosine is added at the end of the extension product. In the second step, this product is extended on a second template. If the terminal A primes where the templates carries a C the result is the incorporation of a deoxyadenosine residue opposite a deoxycytidine residue (taken from Hofreiter *et al.*, 2001)

A second phenomenon, jumping PCR, i.e. the occurrence of template switching during PCR may contribute to these substitutions. It has been reported that lesions in a DNA template can cause the extending primer to jump to another template –either true aDNA or contaminating- during the amplification reaction (Pääbo *et al.*, 1989). When the DNA polymerase encounters the end of a template, it sometimes inserts an adenosine residue, and then jumps to another template creating a chimeric molecule. This can also result in CG→TA substitutions as shown in Figure IV.18.b. Jumping PCR can also lead to hybrids between aDNA and contaminating molecules. These two mechanisms can explain some of the pattern observed in the ancient DNA sequences.

Since these phenomena are unlikely to happen in the same way twice, it is believed that those PCR products which yielded identical sequences indicate genuine aDNA. If the sequences obtained were compromised by the occurrence of deamination in the aDNA it would have been expected to observe differences in sequences among the independent DNA isolations and PCR amplifications from different dental samples (as the examples mentioned).

Quantification of DNA template — although detailed quantification of the DNA template was not carried out, it can be indirectly estimated by the number of PCR cycles (45) that were needed to obtain identifiable PCR products. As seen in Chapter III and according to Rameckers et al. (1997) this will be the number of cycles necessary for an initial template of <1000 molecules.

Amplification from a second extract — for each individual analysed, at least two extracts from two different samples were amplified and sequenced. All the aDNA sequences included in the population analyses yielded identical matching sequences from independent extracts.

Reproduction by a second laboratory — three samples, namely w38, N15 and w49 were replicated by 2 independent labs, yielding DNA sequences that perfectly matched the ones obtained in our lab.

Phylogenetic sense — when DNA from an extinct species is studied its authenticity can be verified by their degree of similarity with orthologous sequences in closely related extant species. This phylogenetic criterion cannot be used in this study to distinguish between real or contaminating sequences, since the most likely source of contamination

is DNA from other humans. However, many other aspects of the aDNA sequences which have been described in detail during this Chapter further supported the authenticity of the aDNA sequences. Summarising, these are: a) the aDNA sequences obtained made clear phylogenetic sense regarding their European origin, as they clustered within the major European haplogroups, as shown by the phylogenetic analyses, b) the observed genetic variability and frequency distribution of haplotypes it is within the range of other human mtDNA data –as opposed to the low haplotype diversity expected as a result of contamination– and c) all the variable sites found in the authenticated aDNA sequences have been previously described as being polymorphic positions in a large survey on mtDNA control region substitutions (Miller *et al.*, 1996) and the frequency distribution is also in agreement with data previously reported (Meyer *et al.*, 1999).

All things considered, the criteria to authenticate aDNA sequences were fully satisfied. Thus, it can be said that the data obtained in this study are authentic HVS-I sequences of 1,700 to 1,000 year old individuals from Anglo-Saxon communities of England and that their subsequent analyses is therefore reliable.

IV.3.2 Genetic diversity

IV.3.2.1 Sample size and estimation of diversity

Several parameters were used to measure and compare the genetic diversity of the ancient and modern populations. At the haplotype level, the ratio of number of haplotypes per sample size (k/N), Nei's heterozygosity, the Shannon index and the haplotypic richness were calculated. As the studied populations ranged from 18 to nearly a thousand individuals in size, special emphasis was given to the effect that difference in sample size may have on the estimation of genetic diversity.

k/N simply shows the proportion of different haplotypes without taking into account any other information, such as the frequency of each haplotype. The number of haplotypes is expected to increase with sample size. However, this is true for small to medium size populations and diminishes while the size of the population increases. This is due to the observed distribution of haplotypes in European populations, with few

moderately frequent lineages and a majority of rare lineages. Given this sort of distribution it follows that the rate of detection of new lineages decreases with sample size (Helgason *et al.*, 2000a; 2000b). This was observed among the modern populations analysed, where, for the best-sampled populations, such as Scotland and Norway, a ~1.6-fold increase in sample size accounted only for an increment of 8% of the haplotypes.

Nei's estimator of gene diversity (here denoted as GD) measures the probability that two alleles or haplotypes sampled at random from a population are different. It incorporates the frequency of haplotypes in the population but is highly sensitive to differences in sample size (Nei, 1987). For example, the Scottish sample has a comparatively high value of gene diversity, despite having a small number of haplotypes per sample size.

The Shannon index is often applied in Ecology to measure species diversity but has also been used in several genetic studies (Bertranpetit *et al.*, 1995; Comas *et al.*, 1997; Varesi *et al.*, 2000). It accounts for both the abundance and evenness of the populations. As this index is also highly dependent on sample size a standardised version was used. Thus, a value of $H'_S = 1$ would imply that all the individuals in a population have different haplotypes. For example, for the abovementioned case of the Scottish sample, despite its large sample size, the corrected Shannon index was comparatively small, more consistent with the actual number of haplotypes. However, this may also be an underestimation of the diversity due to the very large sample size. Very large populations would be expected to give comparatively smaller H'_S indexes, as less new haplotypes will be detected.

The allelic or haplotypic richness also accounts for differences in sample size and was used by several authors as an estimator of diversity for conservation purposes (e.g. ElMousadik & Petit, 1996; Petit *et al.*, 1998; Vendramin *et al.*, 1999; Comps *et al.*, 2001). In this case, the correction for differences in sample size is obtained by the rarefaction method, which uses the data from a larger sample to answer the question of how many species would have been found in a smaller sample of the same population. If n organisms are found in the less-sampled population, the rarefaction method takes hypothetical sub-samples of n organisms from the more-sampled region, and calculates

the average number of species in such sub-samples, allowing comparisons among populations of different sample sizes. This index is purely based on the observed relative abundance of the alleles in the N sampled genes. Originally, according to Sanders (1968), the procedure was to keep the percentage composition of the component species constant but reduce sample size, that is, to artificially create the results that would have been obtained had smaller samples with identical haplotype (faunal in the original paper) composition been taken. However, both Hurberlt (1971) and Simberloff (1972) remarked that the rarefaction method thus defined by Sanders was itself dependent on sample size. This corrected method, proposed by Simberloff (1972) was used in this study. This measure is influenced by the presence of rare variants (alleles or haplotypes) in the population.

In any case, all the estimators showed a similar trend, especially at the extremes of the distributions. The highest values of haplotype diversity were observed for the non-European populations. Among the European populations, the Ladins from north Italy showed the highest haplotype diversity and the lowest diversity was observed for the Saami who are well known for being extreme outlier of Europe for other genetic markers (Cavalli-Sforza *et al.*, 1994; Delghandi *et al.*, 1998; Kaessmann *et al.*, 2002). Of the other classical outliers, the Icelanders, the Basques as well as the Romano-Gypsies showed all low values of gene diversity. This latter population has recently been reported to show particular genetic traits due to be highly endogamous (Gresham *et al.*, 2001; Kalaydjieva *et al.*, 2001). In general, estimations of haplotype diversity based on the 210 bp HVS-I fragment analysed in this study were in agreement with previously reported results based on larger HVS-1 fragments (e.g. Comas, 1996; 1997).

IV.3.2.2 Diversity of Anglo-Saxon communities

The relation between sample size and genetic diversity is well known. All other things being equal, large populations carry more selectively neutral diversity than small populations. The Anglo-Saxon communities analysed are believed to have gathered between 60 and 400 hundreds inhabitants (Evison, 1987; Sherlock & Welch, 1992; Cooper, 1996; Williams & Newman, in prep.; Popescu, pers. comm.). Therefore, due to their relatively small size they were expected to present low haplotype diversity. Whole

families, including men, women and children, must have been buried in the cemetery, and in fact, osteological data suggested some family groups among the skeletal remains excavated (Sherlock & Welch, 1992; Popescu, pers.comm.; Parfitt, pers.comm.). Unlike DNA analyses of modern populations where exclusively non-related individuals are sampled, biological (and maternal) relationships among the individuals of these communities would have accounted for a low genetic diversity. Counter-intuitively, all the diversity indexes showed that the ancient population of England had a high mtDNA haplotype (k/N = 0.75, GD = 0.98 ± 0.07, H'_S = 0.90 and $r_{(48)}$ = 36) and nucleotide (π = 5.51 ± 2.69) diversity ranking always among the top ten populations.

Surprisingly, the genetic diversity observed in these ancient communities was higher than the one observed in the modern population of England both at the haplotype and nucleotide level (k/N = 0.45, GD = 0.93 ± 0.01, H'_S = 0.70 and $r_{(48)}$ = 31.44, and π = 3.81 ± 1.92) (with GD significantly higher, $P = 1.6\ 10^{-6}$). The comparatively higher diversity of the ancient samples does not seem to be due to a sample size effect, as all indexes showed the same pattern. Nor is it likely due to a larger proportion of rare haplotypes, as was indicated by the proportion of unique haplotypes (75%) (or private haplotypes, see Chapter V) which was well within the observed range for other populations ($\bar{x}$ = 72.7 ± 10.9).

High haplotype diversity might have been obtained if the analysed (total) population was sub-structured and its sub-populations highly differentiated. It is believed that the Germanic invaders of England belonged to three different tribes, Angles, Saxon and Jutes (Bede, 731) who inhabited the areas that are now Denmark and north Germany. Each of these tribes is supposed to have disembarked in different parts of the island. Thus, it might well have been the case that the studied Anglo-Saxon sites (located in different areas of England) represented different invader populations. In addition, the Norwich settlement may have been influenced by Norse invaders. The first recorded invasion occurred in 793 AD, at the battle of Lindisfarne. Raids were infrequent up until the middle of the 9th century; however, a significant Viking presence was established on the Northern Islands. Attacks soon followed on British mainland which led to the establishment of settlements in the east coast of England (Keynes, 1997). There is considerable evidence of Scandinavian settlement in northern and midland England.

However, early Scandinavian sites can be distinguished from the Saxon ones only if there are artifacts in the graves. Certain sword types in male burials and the so-called tortoise brooches in female graves would indicate Viking origin (Hall, 1990). This is the case not just because of the specific origin of the finds, but also because late Anglo-Saxon graves are devoid of grave-goods. In fact, very few finds were associated with the settlement of Norwich, dated from the 9th to 11th centuries, only an ansate or equal-armed brooch and two disc brooches (Goddall *et al.*, in prep.). However, nearby the Castle Mall site, a few Viking brooches and honestones possibly from Norway were found (see Fig. II.9) (Penn, 2000). Therefore, it might well have been that the late Anglo-Saxon settlement in Norwich had Viking influence. Again, *if* the Vikings were genetically different then this might also have been reflected in the assessment of the genetic diversity.

However, AMOVA analyses indicated that the diversity *within* the populations always accounted for more than the 92% of the total variation (data not shown). All the settlements (except for Lavington) had high matrilineal diversity and most of them showed diversity values higher than the modern English population. Only Norwich seemed to be genetically differentiated from Norton and Lavington (although F_{ST} values became non significant after correction for multiple comparisons). In fact, a rather different representation of the major mtDNA haplogroups among the communities can be distinguished by the reduced median networks. Norwich had a comparatively larger proportion of hg T and absence of U5 whereas Lavington and Norton had a larger proportion of hg U5. These differences would agree with the F_{ST} values obtained.

The high diversity of the settlements might also be understood as the result of population subdivision *within* the archaeological site –provided the Britons and Saxons were genetically distinct. For the acculturation model's supporters, communities in Anglo-Saxon society would have been composed of a mixture of native people (i.e. Britons) and invaders (i.e. Anglo-Saxons and later on, Vikings) living side by side. Finding of native Britons in Anglo-Saxon cemeteries has been claimed based on different osteological and archaeological data (Härke, 1990; 1995; Jackson, 1995; 1998).

Härke (1990) has suggested that males buried with weapons (about half of all male adults) might be of Germanic origin whilst the group of males without weapons would appear to be rather mixed, but may be predominantly of British origin. In addition, burial position (i.e. the position of the body in the grave) has been also suggested to be indicative of the origin of the interred individual. Thus, prone and crouched burial with a preference for northerly orientation, have been seen as evidence for continuity of British burial costume among Anglo-Saxon communities (Sherlock & Welch, 1992).

In Norton, for example, as north-south alignments are predominant and there are 32 crouched and 7 prone burials, a total of 39 individuals (32.5%) of the cemetery has been claimed to be a potential native British component of the community. If it is correct to associate crouched burial as the rite of an indigenous British element of the population then a larger proportion of the spear-bearing men could be presumed to be descendant of the Germanic settlers (Sherlock & Welch, 1992). Out of the 12 Norton individuals from whom HVS-1 sequences were obtained two, namely samples N9 and N57, were buried in prone/crouched position and three, namely samples N1, N15 and N16, had weapons. Unfortunately the numbers are too small to perform any statistical analyses.

Another way of addressing this issue was to compare the diversity of the ancient communities with that of an artificial population of invaders and natives. Among the invaders, the Saxon component was represented by the samples from north Germany and Denmark and the Viking, by the samples from Norway. Samples from Wales and Cornwall were included as putative representatives of the native Britons. The diversity of this mixed population (k/N = 0.33, GD = 0.94 ± 0.05, H'_S = 0.65, $r_{(18)}$ = 13.92 and π = 4.20 ± 2.09) was lower than the diversity of ancient Britain (k/N = 0.75, GD = 0.98 ± 0.07, H'_S = 0.90, $r_{(18)}$ = 16.18 and π = 5.51 ± 2.69).

At first sight, this result would seem to suggest that the high diversity observed in ancient Britain was therefore not due to a mixture of native and migrant populations. However, it should be noted that the samples used for this comparative analysis are the *modern* counterparts of the putative migrant and native *ancient* populations (i.e. Germany ≈ Saxons, Norway ≈ Vikings, Wales ≈ Britons, etc) and their current observed diversity might actually not reflect their diversity in Anglo-Saxon times. In fact, this seems to be the case for the English population itself; where the current diversity is

lower than the ancient diversity. Therefore, if a gain or loss of diversity over the last millennium had also occurred in any of the modern populations included in the mixed *founder* sample (i.e. north Germany, Denmark, Norway, Wales or Cornwall), this analyses would yield little information regarding the mixed nature of the Anglo-Saxon settlements. In any case, interestingly, the diversity value obtained for the modern mixed population was very similar to that of modern England (k/N = 0.45, GD = 0.93 ± 0.01, H'_S = 0.70 and $r_{(18)}$ = 13.59; π = 4.20 ± 2.69) which would be not surprising *if* the modern population of England is in fact composed of those invaders populations *and if* similar lost of variability occurred in the other modern populations

IV.3.2.3 Expansion times, mutation rates and aDNA analyses

Analyses of mismatch distributions of the modern populations showed that, with the exception of the Saami, all populations showed a uni-modal curve which does not fit the pattern expected for a constant size population. Tajima's *D* values for most of the populations were significantly negative. Although this test evaluates the hypothesis of selective neutrality and equilibrium, negative values can be due to other non-selective factors such as demographic expansions, as reported for most human populations (Rogers & Harpending, 1992; Comas *et al.*, 1997; Jorde *et al.*, 1997; Harpending *et al.*, 1998; Harpending & Rogers, 2000; Simoni *et al.*, 2000; Kaessmann *et al.*, 2002).

According to Rogers (1992) in the case of population growth the resulting curve will travel at rate $2u$ (with u = mutation rate of the fragment analysed and per generation) and its crest will be at $\tau = 2ut$ after t generations. For contemporary populations, mutation rates being equal, mutations would have occurred over time and the longer the time since expansion the more mutations should have accumulated in each population. Thus, the location of the crest provides a rough idea of the time in units of $1/2u$ since the episode of population expansion.

However, the use of mtDNA sequence data to asses the time elapsed since the particular event took place (e.g. expansion or bottleneck times) requires knowledge of the mutation rate in the particular region analysed. Two different approaches have been undertaken to estimate the mutation rate: phylogenetic and pedigree analyses. However, as discussed in the Introduction, no agreement on the mutation rate of the CR has yet

been met. The extensive heterogeneity across the CR of the mtDNA will also affect the estimation of average mutation rates (Howell *et al.*, 1996).

In addition, aDNA presents a particular drawback for the estimation of expansion dates. Given the high degradation of the DNA strands, only very short sequences can be amplified. As mentioned, due to the heterogeneity of the HVS-I, variations in the length of the analysed sequence do not necessarily entail a proportional increase/decrease in the τ values obtained –and the subsequent expansion or bottleneck times (t) calculated. Thus, demographic analysis of a given population by means of a number of HVS-I sequences of different length will not yield similar expansion or bottleneck times (Montiel-Duarte, 2000). In particular, estimations of τ for a short and highly variable HVS-I sequences –as the one analysed in this study– will result in relatively higher values of τ, with a concomitant overestimation of the expansion time (t).

This explains the very early expansion dates obtained in this study both for the ancient and modern populations. Nonetheless, taking the numbers as a rough estimate, these expansion times can still be used for relative comparisons. The results agree with previous analyses of (larger) HVS-I and HVS-II sequences from European and non-European populations. The earliest expansion dates are observed for populations of the Near East and decreasing towards West Europe (Comas *et al.*, 1996; 1998).

The ancient population of England showed a τ value higher than the value obtained for the modern English population. Since the ancient population is shifted to the right, this theoretically should mean that the ancient population expanded earlier than the modern English population. Taking into account that these two samples are allegedly the *same* population sampled at different times (~1,700 to 1,000 apart) this result must be read carefully. Although a large amount of mutation would not have been accumulated in such a short period of time (in evolutionary terms), it would be expected that if anything the modern populations would present more variability than the ancient one. However, the opposite was observed. Thus, the observed mismatch distribution of ancient and modern England would not seem to support the hypothesis of genetic continuity between these two populations, at least not without major demographic or selective events during the elapsed time.

IV.3.4 Phylogenetic analyses

It could be observed by the reduced median networks (RMNs) that not only haplotype and nucleotide diversity has been lost during this last millennium but also there has been a distortion in the distribution of haplotypes and frequencies of haplogroups (P = 0.02). Whereas haplogroup U5 was very common in the ancient communities of England, this haplogroup is not well-represented in modern England. In addition, the lineage of sub-hg J1b, namely [126-145-172-222-261] said to be exclusive to the British population (Richards *et al.*, 1996; 1998) was not observed among the samples of ancient Britain. Haplogroup K, also very frequent in the modern population was less well-represented in the ancient one. Haplogroup T (mainly T1 and T2) was very frequent in the ancient sample (mainly in Norwich) but much less represented in the modern population of England. It might be argued that the absence or low frequency of some of the haplogroups may be due to the small sample size of the ancient English sample, however, the opposite seems unlikely.

Reduced median networks constructed for some of the major clusters (namely JT, IXW and U5) for ancient and modern English haplotypes together indicated that in RMNs of U5 and IXW, the ancient haplotypes were mostly internal nodes of the network. On the other hand, in RMN of JT, the ancient haplotypes were exclusively tips of the branches of the network, in particular of hg T1 and T2. This would indicate that those haplotypes have originated more recently in comparison with the U5 and IXW haplotypes. In addition, these haplotypes were not observed in the modern population. If they had been observed in male individuals, this might suggest they are first generation migrants. However, the individuals showing these haplotypes were exclusively females and children (see Chapter V, Fig. V.10). In any case, whether at that time or later, these hg T haplotypes went extinct.

IV.3.5 Demographic events

The results obtained consistently indicated that the ancient communities that inhabited England during the 4th to 11th century had a high matrilineal diversity, both at the haplotype and nucleotide level. Diversity may be gained either through mutation or through gene flow from neighbouring populations. A millennium is not long enough for

diversity to increase significantly by *de novo* mutations. However, it would have been expected that the several waves of migration into the British Isles during this period and the large increase in population size (from ~0.6 million to the current 60 million; McEvedy & Jones, 1978) would have had a considerable impact, which should have been reflected in an increase of diversity in the modern population of England. Yet, the opposite was observed: matrilineal diversity was higher in Anglo-Saxon communities than in present times. Or in other words, it would seem that genetic diversity has been lost in the last millennium.

Loss of diversity occurs either randomly through genetic drift or actively through selection. The expected loss of variation due to random genetic drift is inversely proportional to the effective size of the population, $H_t = (1-1/2N_e)^t H_0$ (where t is the number of generations and N_e is the effective population size, which for mtDNA is half of the female N_e). Small size populations are consequently more susceptible. The effective size of a population may be altered by several factors, such as unbalanced sex ratio, population substructure (for diploid systems), variance in family size and fluctuating population size (Kimura & Crow, 1963). Thus, the combination of demographic and social factors can contribute to loss of diversity in large populations.

It is possible that the drastic demographic events that took place over the last millennium could have been responsible for the loss of diversity. In particular the 14th century was an era of catastrophes. First, the Great Famine (1315-1317) and then the Black Death (1347-1351). There had been other famines in Europe, but none that had persisted for so long and with such a large population (McEvedy, 1988). It is believed that between 10-15% of the population died, with some places accounting for larger proportions. Later, in 1347, in the middle of another disastrous harvest the Black Death struck. The mortality rate from the first outbreak of the Black Death overall in Europe has been estimated at between the 30% and 45% of the population (McEvedy, 1988).

In England, although the exact size of the population in 1300 and 1348 is not known, it is generally believed that by the end of the 13th century it may have been between 5 and 6 millions, and that after the first wave of the Plague there may have been as few as 2 to 2.5 millions. The population had been nearly halved (Goldberg, 1996).

The main difficulty in estimating the overall rate of mortality resides in the diverse evidence from records from different regions. It is often argued that although there is evidence that some localities suffered heavily during the Black Death, the overall mortality was somewhat lower because other villages escaped almost unharmed (Levett, 1916).

As the population fell in numbers, the demand for food fell and prices were depressed. In contrast, labour costs and wages rose, resulting in a short-term benefit and prosperity for the survivor population. The age of marriage in peasant society, it is argued, is controlled by the availability of land. If it was easy to find a 'niche', a place within the existing socio-economic framework of the village, then marriage would have taken place at a relatively early age. If the economic prospects of the marriage were also good, offering relatively abundant food, adequate housing and clothing then this would also encourage an earlier age of marriage (Bolton, 1996). The earlier the marriage the greater the fertility of the woman within the marriage, and the more children born, which, in the harsh conditions of the Middle Ages, meant a higher chance of some children surviving. Such conditions certainly seem to have been obtained immediately after the first plague.

However, long-term recovery did not occur and the population in England failed to replace itself. The main reason for this, it has been widely believed, was the recurrence of plague. Since 1348 and for many years, plague was pandemic in England, with at least 15 outbreaks during the end of the 14th and 15th century and two major events in 1556-1563 and 1665-7 (Ziegler, 1969; Goldberg, 1996). It was the devastating effect of the subsequent plagues which is thought to have produced the demographic alteration of the population.

Russell (1948) and Heillener (1967) have argued that the recurrent outbreaks of the Plague had both short- and a long-term consequences, the former preventing the recovery, the latter altering the age structure of the population, thus depressing birth rates in succeeding decades. Some figures implied that during the later 14th and 15th centuries few children were produced per marriage and sometime no children at all. Replacement rate (the ratio of inheriting sons to deceased fathers) fell by 50% between 1351 and 1375 both in peasant and noble society (Campbell, 1984; Kowaleski, 1984).

In summary, it would seem that those who stayed where they were and worked the land, the agriculturalists, married earlier, but significant numbers of labourers, perhaps up to a half of the population married much later, if at all (Goldberg, 1988).

Other epidemics, such as the sweating sickness, in 1485, may have contributed to the slow recovery of the English population. Unlike the Plague, this fatal disease, characterised by sudden sweat and a pulmonary condition, was specific to England. The epidemic was widely scattered and patchy in its geographic distribution, with a quick onset. Death from the disease was estimated to have affected 20% of the population (Thwaites *et al.*, 1997). Moreover, in following centuries the population was also affected by other large epidemics, both worldwide epidemics (influenza in 1723, 1775, 1857 and 1889; cholera in 1826-1837 and 1848 and Spanish influenza in 1917) and those restricted to England (cholera in 1853; smallpox in 1947) (Creighton, 1894).

Thus, it might be possible that in addition to the well-known devastating effect of the Black Death and other epidemics, the resultant demographic changes in the population have had consequences so far not realised. As mentioned, it would seem that birth rate and family size increased for some of the population whereas for others the opposite was the case. If so, this would result in a large variance in family size, with some individuals contributing disproportionately more to the next generation than others. Even for large populations this would result in a concomitant loss of diversity.

If so, it might be expected that other European populations would have gone through a similar process. Ancient DNA analyses of European populations are very scarce (e.g. Izagirre & de la Rúa, 1999; Cipollaro *et al.,* 1998) so that comparisons are not possible. Gerstenberger (2002) analysed an early medieval population in Germany in order determine wedding patterns. Unfortunately, a shorter and less informative HVS-I fragment was analysed so that no comparisons can be made.

In addition, another factor may have also enhanced this effect. It is known that whereas whole families and even villages were wiped out by the Plague, other persons would seem to be immune or resistant to it. A well-known case is that of the population of Eyam, known as the Plague village. When the third strike of the Great Plague, got to Eyam in 1665 the population, under the directions of the rector, locked themselves in the village so that the Plague would not spread to other unaffected neighbour villages.

As a consequence by the end of 1667 two thirds of the population died, yet a third, which was certainly exposed to plague, survived (Clifford, 1989). Genetic analyses carried out in descendants of Eyam's surviving individuals indicated that a large proportion of them shared a mutation in a cell receptor gene (CCR5) (O'Brien, see Uhl, 2001). The mutation consists of a 32-bp deletion, resulting in a truncated receptor which is not expressed in the cell membrane (Dean *et al.*, 1996). It has been argued that without this receptor the bacillus causative of the Plague, *Yersinia pestis,* might not enter the cell and therefore could not infect it. Thus, carriers of this mutation (named Δccr5) would be resistant –or much less susceptible– to Plague (O'Brien, see Uhl, 2001). In fact, this finding was a by-product of investigation on the chemokine receptor gene CCR5 which was already the object of interest due to its role in the entry of HIV-1 into target cells (Dean *et al.*, 1996). The same described mutant allele, Δccr5, which results in absence of the cell receptor, confers resistance to HIV in homozygote individuals (Samson *et al.*, 1996). The frequency distribution of Δccr5 and its strong linkage disequilibrium with specific alleles of microsatellite loci, suggested a single and recent origin for the mutation (Martinson *et al.*, 1997; Libert *et al.*, 1998; Stephens *et al.*, 1998), with an estimated age between 700 to 3,500 year old (Libert *et al.*, 1998; Stephens *et al.*, 1998). It has been postulated then that, the high frequency of Δccr5 among Europeans is unlikely to be caused by simply effect of genetic drift (Libert *et al.*, 1998). Instead, selective pressure associated with resistance to an infective disease may account for this frequency and distribution. It was thus postulated the Plague might have acted as a positive selection pressure for the mutated gene (Stephens *et al.*, 1998).

It has recently been argued that *Yersinia pestis* might have actually not been the causative agent of the Black Death (Paterson, 2002; Reese, 2002) based on mortality rates obtained from analyses of bishop's records of the replacement of priests (Wood *et al.*, 2002a; 2002b) as well as the spreading pattern of the epidemic. However, other studies have definitely found genes of *Yersinia pestis* in individuals who died during Black Death (Raoult *et al.*, 2000; Drancourt & Raoult, 2002) confirming its being the causative agent of the Plague and Black Death.

The inheritable resistance to plague together with the pandemic condition of the disease would have meant that those individuals that survived the first outbreak would

have likely survived further outbreaks, and that their children would also have a greater chance of survival. This might have enhanced the disproportional contribution of some individuals of the population to the next generation, who, in due time (i.e. by the following strike of the Plague) would have also contributed more to the following generation. In this manner, each strike of the Plague would have acted as a sieve, letting go through the same related individuals (i.e. genes) hence enriching the population with those lineages that survived. It might be expected then that other genetic characteristics of the population would have been enriched as well. It should be noted that this does *not* require linkage between Δccr5 and mtDNA haplotypes.

All things considered, it would seem that the overall effect of the Plague (i.e. an increase in family size variance augmented by the inheritable resistance and repeated outbreaks of the disease) has produced elimination –by drift– of some mtDNA variants and increase of others. This resulted in a concomitant change in these haplotype relative representation and diminution of diversity in the modern population of England. As argued earlier, it would be expected that other modern populations might have gone through a similar demographic process. Only comparative ancient DNA analyses as in this study would allow confirmation of this hypothesis. If true, conclusions drawn from the analyses of the current European genetic diversity with regards to prehistoric processes should be re-assessed.

Chapter V - Geographical origin of mtDNA of ancient England

V.1 INTRODUCTION

When people move they take their genes along and pass them on to their descendants in their new homes. Thus, populations retain clues to their origins and roots. Genetic analyses offer the possibility of understanding movements and events of human history.

V.1.1 Phylogeography and founder analyses

The study of the geographic distribution and diversity of genetic variation, known as the "phylogeographic approach" is a useful tool for the investigation of human prehistoric expansions and migrations. It is particularly suited for the study of non-recombining markers, such as mtDNA, which, as it is inherited down the female line and evolves rapidly, allowing a good construction of female genealogies (Avise *et al.*, 1987). In addition, it has a smaller effective population size than nuclear loci, rendering it more sensitive to founder effects and genetic drift –a useful property for markers used to detect geographical substructure.

In one of the first large-scale mtDNA analyses of European populations, Sajantila *et al.* (1995) attempted to detect evidence for geographic structuring of mtDNA variation and correlation with language distribution. The study concluded that the capacity of mtDNA to reveal geographic patterns of genetic relationships between European populations was unpromising, as most of the observed diversity was found within populations rather than between populations. However, Richards et al. (1996;1998) proposed that a phylogenetic analyses by RMNs was more appropriate to identify any relationship between DNA sequences and populations (see Introduction). In this way they could detect that some of mtDNA haplogroups were restricted to certain geographical regions (Richards *et al.*, 1996; 1998). Interestingly, one of these sub-clusters was restricted to Britain. This was a sub-group of J1 (originally group 2A)

which they referred to as 2A-w given that was only observed in populations of western Europe, and presents the common segregating sites [126-145-172-222-261]. It might be therefore possible to use this haplotype and its derivates as mtDNA marker for British individuals in admixed populations.

A particular phylogeographic approach known as founder analyses has also been used to investigate the movements of peoples. It is based on a model of population history that sees new populations emerging as the result of older populations throwing off occasional small groups of colonist who move away and establish a new community elsewhere (Richards & Macaulay, 2000). The aim is to identify migration events from a putative source population to a putative descendant population. The source population may be identified, for example, on the basis of archaeological or historical evidence. Several previous studies have applied the basic principle of mtDNA founder analyses to investigate migration and colonization processes, as well as their timing, of both prehistoric (Torroni *et al.*, 1992; 1993; Sykes *et al.*, 1995; Forster *et al.*, 1996; Richards *et al.*, 1998) and more recent human history (Helgason *et al.*, 2001).

However, a different application of founder analyses can help understand colonization events where the source population(s) remains unknown. An example of this is the study by Helgason et al. (2001) in the population of the North Atlantic islands. During the late 8th century Vikings colonised the islands of the North Atlantic in their westward spread. Whereas Orkney and the Western Islands are know to have had thriving pre-Viking settlements of Picts and Gaels, respectively, Shetland was less densely populated and the Faores Is. and Iceland are both believed to have been largely uninhabited at the time of the first Viking settlements. Iceland is still inhabited by descendants of those settlers, some of whom are though to have originated from the British Isles (Helgason *et al.*, 2000), but the Viking legacy is less clear for Orkney and the other smaller islands. Helgason et al. (2001) attempted to assess the relative proportion of Gaelic and Scandinavian admixture in the mtDNA gene pool of the North Atlantic island populations by founder analyses. Thus, genetic distances, namely ρ distances, were calculated between the populations of the North Atlantic and putative source populations (Norway, Scotland, etc) in order to identify the original source(s) and the extent of admixture. It is doubted that any legitimate interpretation should be drawn

from the results obtained, as the authors disregard the effect of sample size in their computations (see this Chapter for details), however, the basic rationale of this analysis is entirely adequate and –corrected– may allow the identification of source populations of other migration events.

V.1.2 Anglo-Saxons in Britain

Since the early times Britain went through several periods of cultural change. Following the depopulation during the LGM and the subsequent resettlement by Mesolithic hunter-gatherers around 10,000 YBP (Dyer, 1990) the Neolithic transition associated with the beginning of farming communities took place. It followed the 'arrival' of Bronze-Iron Age and Celtic culture, and later on the several waves of occupation, such as Romans, Saxons, Vikings and Normans. However, cultural transitions before 0 AD, in particular the Celt culture, are no longer interpreted as implying migration of Celtic tribes from a Celtic homeland (Chapman, 1992; Simon & Rigby, 1997). Modern archaeologists emphasise that inhabitants of the British Isles, i.e. Britons, shared as many similarities as differences with continental Celts and are therefore not believed to represent the same ethnic group (Simon & Rigby, 1997). For cultural transitions after 0 AD, the extent to which they actually involved population movement varies in each case, with the Roman invasion being an example of elite dominance (Arnold, 1984). In any case, the most hotly debated of all these cultural transitions is the role of the Anglo-Saxon migration in the sudden change from Romano-British to Anglo-Saxon Britain (Hamerow, 1997). It is widely accepted that in the early years of the fifth century, the withdrawal of the Roman garrisons took place. Although Britain still remained part of the Roman world, the predominance passed from the Romanic to more Celtic elements in the populations (Hunter Blair, 1956). By this time, Germanic peoples (probably from current north Germany and Denmark) who had long been pirates to the inhabitants of Britain altered their intentions towards the island and began to establish settlements. Thus, in the middle of the fifth century a firm foothold had been established along the eastern shores, spawning important secondary settlements (Loyn, 1984). However, neither the historical nor the archaeological data seem to be sufficient to explain the nature and extent of the Anglo-Saxon invasion.

It is though that these settlers were of heterogeneous origin. According to Bede (731 AD) –whose Ecclesiastical History of the English People is the main narrative of this period– they came from three 'races' of Germans: the Saxons, the Angles and the Jutes. Studies on the provenance of the first newcomers were based on associated grave-goods (particularly jewellery), method of burial and distribution of place-names. Intensive work showed a marked difference in the distribution of types of brooches, which was thought to indicate regional peculiarities that may stem from a variation in racial origin (Leeds, 1945). However, the distribution of social costumes and place-names revealed a general measure of uniformity, suggesting that the tribal kingdoms were actually established in England and determined by political boundaries rather than by sharp distinctions of racial origin (Yorke, 1990).

Germanic immigrants might have come in substantial numbers and were apparently used to a stratified society (Loyn, 1984). However, the nature and extent of the settlement is still under discussion. On one hand, it is thought that the Anglo-Saxons came in great numbers with their families, exterminating, or at best driving westwards, the natives whom they met in the course of the migration (Leeds, 1945; Myres, 1986). On the other, that they were few in number, an elite of aristocratic warriors, imposing upon a large population of slaves their language, institutions and costumes (Arnold, 1984). Hence, the extent of the invasion may be seen as whether or not the invaders brought their families and wives with them.

Regarding the male invaders, genetic analyses showed evidence of a mass migration of Anglo-Saxon men into Britain. Comparative Y-chromosome analysis of English, Welsh and Frisian (male) populations indicated that whereas the English and Welsh samples were significantly different, the English and Frisian samples could not be differentiated (Weale *et al.*, 2002). However, no clear signal of Anglo-Saxon women has so far been detected in the mtDNA pool of modern England. Given that the modern English mtDNA seems to be a limited representation of all the genetic variability present ~1,500 years ago (see Chapter IV), only the direct analysis of Anglo-Saxon peoples will allow addressing that question.

V.1.3 Aims

This chapter will investigate the geographical origin of the female component of the populations of Anglo-Saxon England. This will be carried out by looking at the genetic differences and/or similarities among the mtDNA pools of the ancient and modern European populations. For this, genetic distances between populations will be calculated based on nucleotide sequence and haplogroup frequencies. In addition, haplotype sharing and founder analyses will allow a more detail investigation of the possible origin of the matrilines present in the Anglo-Saxon England. Relationships between populations will be also depicted by reduced median networks.

V.2 RESULTS

V.2.1 Genetic differentiation

V.2.1.1 F_{ST} statistics

As a first attempt to identify genetic differentiation (or association) among the ancient and the modern samples, F_{ST} statistics based on nucleotide sequences were calculated. Estimators for the 51 populations are shown in Appendix D. Overall, F_{ST} statistics showed very little differentiation among populations. Only the Saami and the Gypsies were consistently significantly different ($P < 10^{-6}$) from all the others. The ancient population of England showed significant F_{ST} values ($P < 10^{-6}$) only with the Saami, the Gypsies and also with the Basques. This latter are considered to be a relict sample from Paleolithic populations (Bertranpetit & Cavalli-Sforza, 1991; Bertranpetit *et al.*, 1995; Calafell *et al.*, 1996; Wilson *et al.*, 2001). Figure V.1 shows the F_{ST} matrix represented in a two-dimensional space by means of a multidimensional scaling (MDS) analysis.

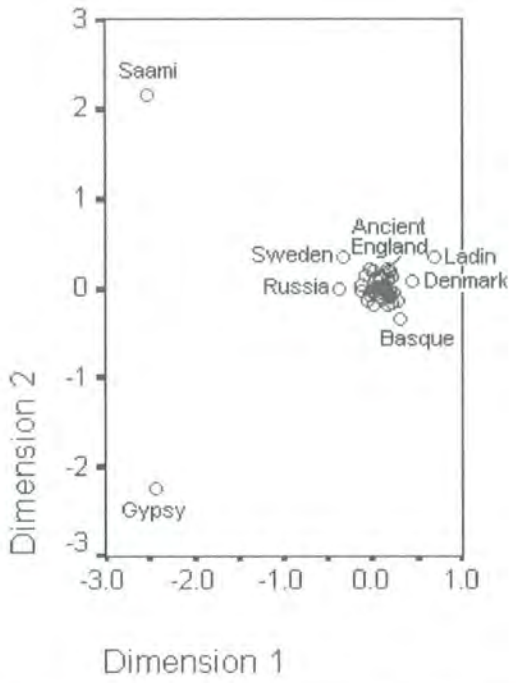


Figure V.1 Multidimensional Scaling Plot for F_{ST} statistics. The 51 dimensions of the genetic distance matrix were reduced to two dimensions, which account for 89% of the variation defined by the original distance matrix. Only outlier populations are labelled and the ancient population of England is further indicated.

It can be visualised graphically that the majority of the populations formed a dense cluster with only the Saami and Gypsy populations being differentiated from the rest. A few populations, including ancient England, Denmark, the Basques, the Ladins (north Italy) the Uralic region of Russia and Sweden were located in the periphery of the cluster.

Given the small F_{ST} values and in order to investigate the overall divergence of the populations, *average* F_{ST} values (referred to as $\overline{F}_{ST}$) were calculated. For this, for example for Austria, its average F_{ST} value was obtained by adding the F_{ST} values of Austria-Belarus, Austria-Romania, Austria-Turkey and so on and dividing the total number of paired population (50). These $\overline{F}_{ST}$ values are shown in Table V.1.

Population	F_{ST}
Gypsy	0.173
Saami	0.173
north Italy	0.042
Sweden	0.032
Uralic Russia	0.029
Basque Country	0.029
Denmark	0.026
north Germany	0.021
Iceland	0.021
Switzerland	0.020
Is.skye	0.020
Karelia	0.020
Wales	0.020
Rome	0.020
ancient England	**0.020**
Western Is.	0.019
Finland	0.019
Turkey	0.018
north Ossetia	0.018
Adygei	0.018
Norway	0.018
Palestine	0.018
Scotland	0.016
south Germany	0.016
Spain	0.016
Georgia	0.016
Greece	0.016
Ireland	0.016
France	0.015
Romania	0.015
Orkney	0.015
European Russia	0.015
west Siberia	0.015
Sicily	0.015
Estonia	0.014
Portugal	0.014
central Germany	0.014
Belarus	0.014
England	**0.014**
Austria	0.013
Armenia	0.013
Sardinia	0.013
Cornwall	0.013
Bulgaria	0.012
Poland	0.012
Czech Rep.	0.011
Bavaria	0.009
Tuscany	0.009
Syria	0.008
Volga-Finn region	0.000
Ukraine	-0.001

Table V.1 Average F_{ST} values ($\overline{F}_{ST}$) for all 51 populations analysed, shown in decreasing order. Values for ancient and modern England are indicated in bold type.

The Saami and Gypsy populations showed identical $\overline{F}_{ST}$ values, which were four times larger than the next population, north Italy. The Anglo-Saxon population ($\overline{F}_{ST}$ = 0.020) ranked among the top ten populations together with Northern and Scandinavian populations, as Uralic region of Russia, Sweden, North Germany, Denmark and Iceland, and also the Isle of Skye, Switzerland and the Basques, some of which had been previously identified as outliers in the MDS graph. The modern English population showed a lower $\overline{F}_{ST}$ value ($\overline{F}_{ST}$ = 0.014), which in principle was not significantly different from the ancient population (P = 0.3). However, when the Saami and Gypsies are removed from the analyses the difference became significant (P = 0.0014).

V.2.1.2 Genetic distances – nucleotide sequences

Nei's genetic distances between populations were calculated based on nucleotide sequences and assuming the Tamura and Nei model of nucleotide substitution with α = 0.26 (Meyer *et al.*, 1999). Two genetic distances were computed: d_{XY} and also the net (or corrected) d_A, which subtracts average diversity within populations from the between-populations component ($d_A = d_{XY} - (d_X + d_Y)/2$). Very little differentiation was found for both d_{XY} and d_A distances, with very small values of d_A (see Appendix D). As for the F_{ST} statistic, only the Saami and Gypsy populations were consistently differentiated from all the rest ($P < 10^{-6}$). The ancient sample was significantly different from these two populations and also from the Basques ($P < 10^{-6}$). In general, pairs of populations that showed significant differences for the F_{ST} statistic also did for d_A genetic distances

Figure V.2 shows the d_A matrix represented in a two-dimensional space by means of a multidimensional scaling (MDS) analysis. The result is similar to the F_{ST} statistic, with the majority of the populations clustering in a compact cluster and the Saami and Gypsy well separated from it.

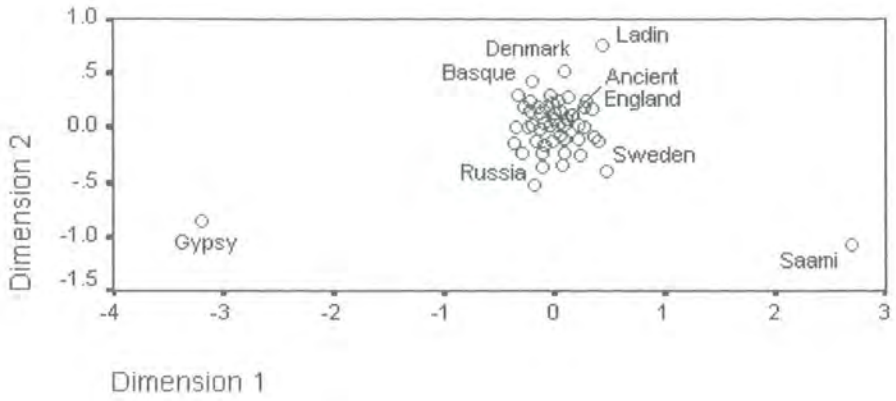


Figure V.2 Multidimensional Scaling Plot for d_A genetic distance matrix. The 51 dimensions of the genetic distance matrix were reduced to two dimensions, which account for 97% of the variation defined by the original distance matrix. Only outlier populations are labelled and the ancient population of England further indicated.

As for the F_{ST} statistics, average genetic distances were calculated for each population. Average d_{XY} and d_A values ($\overline{d}_{XY}$ and $\overline{d}_A$) are shown in Table V.2. The Gypsy population showed the highest average distance value for both $\overline{d}_{XY}$ and $\overline{d}_A$. Two populations from the Near East, two Italian populations and ancient population of England ($\overline{d}_{XY} = 5.056$), also showed very high genetic distance values for average $\overline{d}_{XY}$. On the other hand, modern England showed a lower value ($\overline{d}_{XY} = 4.241$), significantly different from the ancient population ($P < 10^{-6}$). The population showing the smallest distance to the other populations was Poland, followed by the Basques and the populations of Celtic origin of Wales, Ireland and Cornwall. However, the pattern was different for $\overline{d}_A$. Differences between relative position of $\overline{d}_{XY}$ and net value $\overline{d}_A$ for each population are due to the population's internal diversity, d_X. Thus, for $\overline{d}_A$, the Basques (which have a very small d_X. of 2.85, see Appendix D) were among the populations with the highest genetic distance values ($\overline{d}_A = 0.111$). On the other hand, the sample from Rome had a comparatively lower net genetic distance ($\overline{d}_A = 0.093$), due to its high d_X value of 5.743. In the same manner, the ancient sample of England had a $\overline{d}_A$ value of 0.094 which, although still comparatively high, was lower than its respective average for d_{XY}). The population with the smallest average net distance value was Ukraine ($\overline{d}_{XY} = 0.04$).

Population	$\overline{d}_{XY}$
Gypsy	5.713
north Italy	5.320
Palestina	5.219
Rome	5.125
Turkey	5.064
ancient Britain	**5.056**
Saami	5.006
north Ossetia	4.987
Armenia	4.929
Belarus	4.889
European Russia	4.844
north Germany	4.823
Syria	4.810
Denmark	4.759
Uralic Russia	4.755
Tuscany	4.720
Georgia	4.629
Iceland	4.566
Adygei	4.564
Is.Skye	4.498
Volga-Finn region	4.482
Sweden	4.481
Austria	4.434
Ukraine	4.420
Sardinia	4.418
Western Is.	4.382
Orkney	4.366
Czech	4.346
Romania	4.344
Estonia	4.326
Scotland	4.325
Finland	4.249
England	**4.241**
Bulgaria	4.237
Bavaria	4.222
west Siberia	4.190
Spain	4.155
Norway	4.143
France	4.123
Greece	4.122
central Germany	4.100
Karelia	4.084
Portugal	4.023
south Germany	3.999
Switzerland	3.985
Sicily	3.955
Wales	3.924
Ireland	3.923
Cornwall	3.902
Basque Country	3.726
Poland	3.661

Population	$\overline{d}_{A}$
Gypsy	0.991
Saami	0.854
north Italy	0.224
Romania	0.153
Sweden	0.148
Uralic	0.137
Denmark	0.127
Basque Country	0.111
north Germany	0.105
Iceland	0.103
Palestina	0.100
Ossetia	0.097
Turkey	0.095
ancient Britain	**0.094**
Is.skye	0.094
Rome	0.093
Western Is.	0.089
Switzerland	0.089
Karelia	0.087
Adygei	0.086
Wales	0.084
Finland	0.084
Georgia	0.083
Volga-Finn region	0.080
Armenia	0.076
Scotland	0.075
Norway	0.074
France	0.071
Orkney	0.071
Spain	0.071
Greece	0.070
west Siberia	0.070
European Russia	0.069
Portugal	0.069
Estonia	0.068
Poland	0.068
Belarus	0.067
Austria	0.067
south Germany	0.067
Ireland	0.067
Sicily	0.063
central Germany	0.063
Sardinia	0.063
England	**0.061**
Bulgaria	0.057
Cornwall	0.055
Czech Rep.	0.055
Bavaria	0.046
Tuscany	0.043
Syria	0.043
Ukraine	-0.001

Table V.2 Average d_{XY} and d_A genetic distances (derived from Nei, 1987), shown in decreasing order. Values for ancient and modern England are shown in bold type.

V.2.1.3 Geometric distances – haplogroup frequencies.

Genetic distances among populations were also calculated based on haplogroup frequencies. Haplogroup assignment proceeded according to the algorithm based on presence of the relevant segregating sites (see Methods II.3.3.3). On these bases, haplogroup and sub-haplogroup frequencies were calculated for all populations (see Appendix E). However, it should be noted that segregating sites 343 and 356 which define hg U3 and hg U4, respectively, were not available for analyses, as they are beyond the end of the sequence analysed. Thus, sequences which might have been assigned to hg U3 or hg U4 (had those positions been known), have been included into hg H'. Therefore hg H definition did not entail a strict phylogenetic criterion (it is not monophyletic) and hence, the use of nomenclature hg H' rather than hg H.

Genetic distances between populations based on sub-cluster frequencies were calculated using the *f* distance, which is based on the geometric chord distance described by Cavalli-Sforza and Edwards (1967). The resulting matrix of genetic distances between populations, represented in a two-dimensional space by means of a multidimensional scaling (MDS) analyses is shown in Figure V.3.

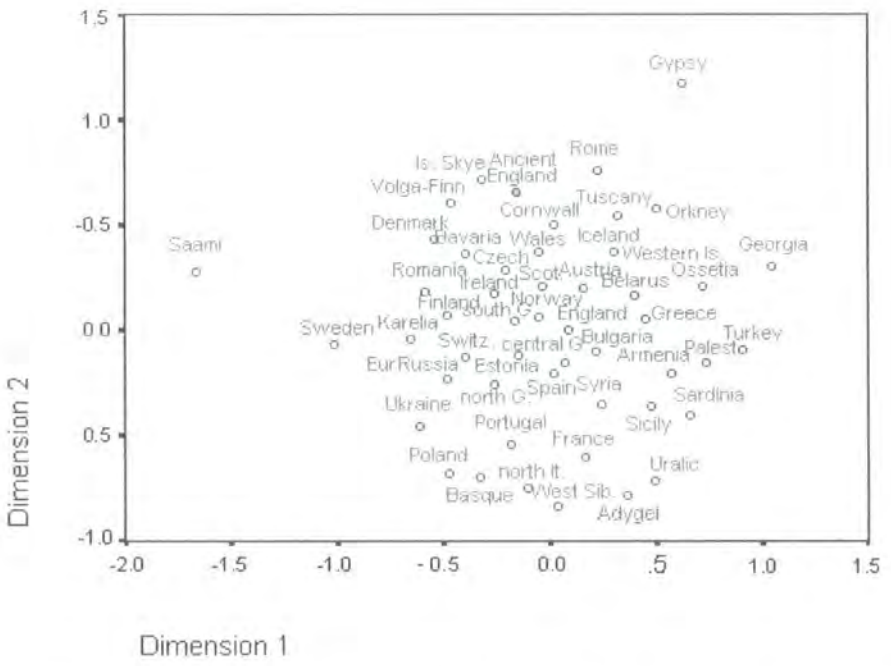


Figure V.3 Multidimensional Scaling Plot of *f* distances (Cavalli-Sforza & Edwards, 1967) based on haplogroup frequencies. The 51 dimensions of the genetic distance matrix were reduced to two dimensions, which account for 91% of the variation defined by the original distance matrix

Although this graph also showed a single large cluster of populations, it seemed that analysis of haplogroup frequency allowed a better resolution. The Saami and Gypsy were clearly differentiated but this time all populations could be singled out as well. Most outlier populations were as seen in previous analyses. Figure V.4 shows a MDS for the genetic distance matrix for *f* distances calculated as previously, but with the ancient population divided according to their historical period (as done in Chapter IV). As for the ancient sample considered as a whole, both the early and late Anglo-Saxon archaeological sites appeared at the periphery of the cluster of populations. Interestingly, these two groupings were at opposite sites of the diagram.

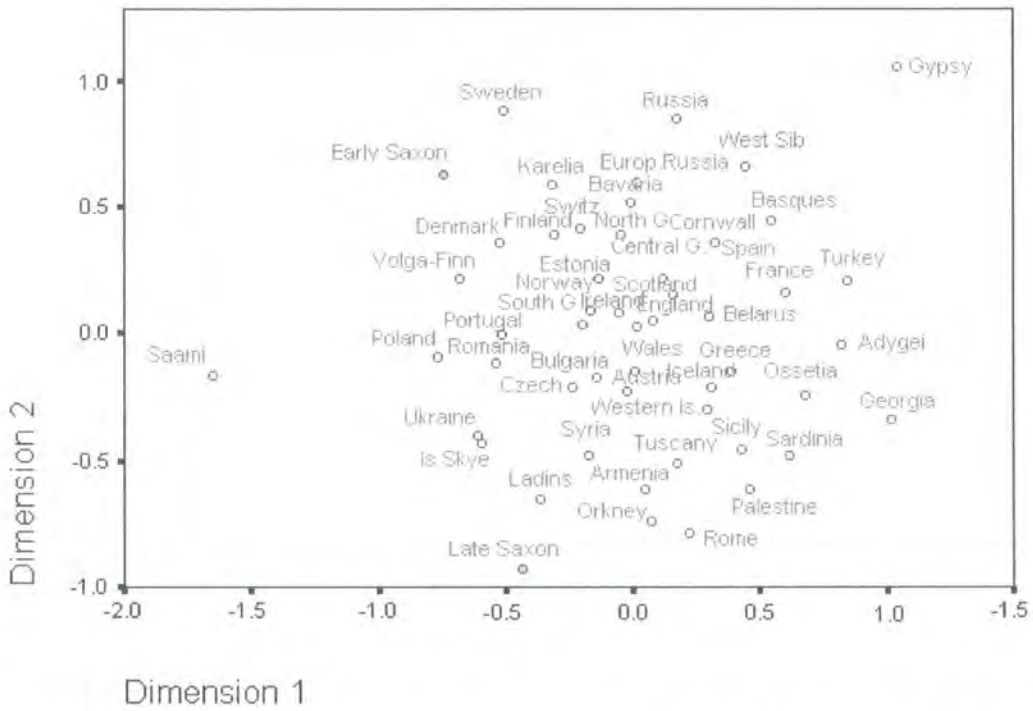


Figure V.4 Multidimensional Scaling Plot of *f* distances (Cavalli Sforza & Edwards, 1967) based on haplogroup frequencies. The 52 dimensions of the genetic distance matrix were reduced to two dimensions, which account for 90% of the variation defined by the original distance matrix. The ancient population was divided into two groups according to their historical period: *Early Saxon* included the settlements of Norton, Buckland and Lavington (5th-7th century) and *late Saxon* included the settlement of Norwich (9th-11th century).

The early Saxon archaeological sites showed the closest association within this cluster to the northernmost populations of Sweden, Denmark, Finland, Karelia and European Russia and to a lesser extent to the Volga-Finn region, Norway, Estonia,

Bavaria, north Germany and Switzerland. On the other hand, the late Saxon site showed closer association with Orkney, the Is. of Skye and the Ladin population of north Italy. Lesser associations were also seen with the easternmost populations of Palestine, Syria, Armenia, Georgia, Ossetia and Adygei, and also the Italian samples from Tuscany, Sardinia and Sicily. The sample from modern England showed an intermediate distance to the two ancient sub-groups, plotting in the middle of the graph. Interestingly, it clustered with Ireland, Scotland and Wales. Cornwall was also relatively close to them, grouping with France, Spain and the Basques. To the left of the graph another group of neighbour populations plotted together: Poland, Ukraine, Romania, Bulgaria and Czech Republic. Intriguingly, Portugal also appeared next to these populations, as did the Isle of Skye.

V.2.2 Haplotype distribution

Another way to assess differentiation and/or relationship among populations is to calculate their proportion of private and shared haplotypes. For this, samples from all 51 populations were pooled together and analysed using Arlequin 2.0 in order to identify the identical haplotypes shared among populations. Of the total 6454 HVS-I sequences compiled for this study, 1412 different haplotypes were found. Of these, 939 were unique, i.e. occurring only once and, therefore, in only one population. These will be referred to as *unique private* haplotypes. Of the 473 multiple haplotypes (observed two or more times in the database), 358 were shared between 2 or more populations. The remaining 115 multiple haplotypes were observed only in one population each, i.e they occurred multiple times but just in one population. These will be referred to as *multiple private* haplotypes. The raw data showing the haplotype distribution of the whole database can be seen in Appendix F.

V.2.2.1 Private haplotypes

The proportion of private lineages ranged widely from 5% to 66%, with an average of $\bar{x} = 32.8$ ($\pm$ 10.5). Only 28% of the haplotypes of the ancient population of Britain were private haplotypes, suggesting that their comparatively high genetic differentiation

is not due to an excess of private lineages. On the other hand, the modern population of England showed a higher proportion of private haplotypes (34.4%).

The population showing the highest proportion of private lineages was North Italy (66%). Rome, the second highest population, had a much lower percentage private haplotypes (50%), followed by eastern populations, such as Palestine (49.5%), Armenia (46.9%) and Georgia (45.3%) and the northern populations of north Germany (44.7%), Norway (43.5%), the Saami (43.35) and European Russia (42.8%). The population with the smallest proportion of private haplotypes was Poland (5%), followed by Ukraine (14.3%), North Atlantic islands of Skye (18.2%) and Orkney (18.4%) and the Caucasian population of Adygei (18.5%) (see Appendix F).

In any case, the fact that the ancient population of Britain has a 28% of mtDNA haplotypes which have not been observed anywhere, suggests that these haplotypes have gone extinct.

If only *multiple private* haplotypes are considered (which were observed in 62% of the populations) their proportion ranged between 14% and less than 1% (average of 5.2 ± 3.7). Only 2.7% (one haplotype) of the ancient English samples were *multiple private* haplotypes. The modern English samples had an even lower value of 0.8%, the smallest value of all populations. The population showing the largest proportion of multiple private haplotypes were the Gypsies (13.9%), followed by Iceland (12.7%), North Ossetia (12.5%) and Finland (12.1%). These results were in agreement with previous analyses of the Gypsy population (Gresham *et al.*, 2001; Kalaydjieva *et al.*, 2001) and the Icelandic (Helgason *et al.*, 2000) and Finish populations (Lahermo *et al.*, 1996; Kaessmann *et al.*, 2002).

The *multiple private* haplotype observed in the ancient population of Britain presented segregating sites [162-292] (haplotype number 432 in Table V.3). It occurred twice in the early Saxon settlement of Buckland, strongly suggesting biological relationship between these two individuals. Based on archaeological data, the individuals are two males of age 25-30 and 35-40 (at the time of death) (see Materials II.1.1.3 and Fig.V.10), possibly siblings on the basis of their age and shared haplotype.

V.2.2.2 Haplotype sharing and phylogeography

The CRS was by far the most common haplotype accounting for ~18% of the data. It was the only haplotype that occurred in all 51 studied populations. The next most common haplotype, namely [126], was observed in 44 populations (86%) and accounted for 4% of the data. Other common haplotypes were those defining the major haplogroups and sub-haplogroups, which were observed in 20-40% of the populations. The remaining ~1400 HVS-I haplotypes occurred at a frequency lower than 1%.

Some geographical pattern could be observed for a few haplogroups. Haplotypes belonging to haplogroup U5 were most commonly observed in north European populations and sub-haplogroup U5b1 was almost exclusive to the Saami (see haplotypes number 32, 354 and 851-3 in Appendix F). Three east Eurasian hg Z motifs were observed exclusively in Nordic populations (see haplotypes number 86, 805 and 856 in Appendix F). Haplogroup U1 was exclusive of near Eastern and Mediterranean populations and hg U6 was only found in Palestine and the Iberic peninsula. Sub-haplogroup K1 was observed in the British and North Atlantic Isles, Iceland and Norway and also in two Balkan populations (see haplotypes number 34, 413, 602, 798, 802, 1377 in Appendix F). Sub-haplogroup J1a was observed in central and northern Europe and also in Near East (although not the same haplotypes). Sub-haplogroup J1b was found almost only in the Near East (see haplotypes number 88, 116, 271, 436, 999, 1118-9 in Appendix F). A smaller group of J1b, namely sub-haplogroup J1b1, was exclusively from Britain and Scandinavia. It had previously been reported that this motif was exclusive to Britain (Richards *et al.,* 1996; 1998). However, the large data set compiled in my study pointed to the fact that such mtDNA motif is actually shared exclusively between Britain and Scandinavia (see haplotypes number 17, 33, 149, 922, 1050 and 1078 in Appendix F).

Special attention was given to those haplotypes which were shared among just a few populations. More often than not, pairs of populations which shared haplotypes were geographically close to each other, such us Scotland-Western Is., Austria-north Italy and Estonia-Karelia, for haplotypes [126-242], [129c-183c-209-260], [134-153-172], respectively, among others. A similar pattern was observed for those haplotypes shared among three or four populations such as, Scotland-Cornwall-Western Is. or Georgia-

Adygei-north Ossetia-Gypsies, for haplotypes [224-311-319] and [126-193-256], respectively.

The haplotype sharing pattern of the ancient samples was studied in detail and is shown in Table V.3. Of the 36 HVS-I haplotypes, ~28% were exclusive to the ancient population, 39% were shared with 2 to 20 modern populations and 33% were shared with more than 20 modern populations. Among the latter, one haplotype, the CRS was, as already mentioned, shared with all the modern populations. Haplotypes [126] and [311] were also very common and shared with other 43 modern populations. Other common haplotypes defining the major hg and sub-hg were also distributed world-wide.

However, for a few relatively common haplotypes some pattern of geographical distribution was detected. For example, haplotype [192-256-270-291] (number 40 in Table V.3) was observed in samples belonging to North Europe and Scandinavia (Germany, Finland, Iceland, Norway, Estonia and European Russia) and the British Isles (Wales, Western Is., Scotland, Orkney and modern England). Another haplotype [221] (number 42 in Table V.3) was shared between these latter insular populations and central-south European populations. The ancient haplotype [192-256-270-311] was also observed in populations from the Caucasus area (Georgia, Adygei and north Ossetia), Norway and some of the North Atlantic Islands (Orkney, Is. Skye and the Western Is.). In ancient England this haplotype was found in two male individuals from the settlement of Norton, of age 35-45 and 25-35, possibly another pair of siblings. Interestingly, both individuals were sexed –based on osteological and anatomical measures– as male, yet their gender –based on archaeological grave-finds– was as females.

Of the 10 ancient HVS-I haplotypes which were shared with less than ten populations (28%), 5 haplotypes were shared with other 6, 5, 4, 3 and 2 modern populations (respectively). Thus, haplotype [126-153-294] occurred in the ancient population of England, Austria, central Germany, Norway, Iceland, the Western Is. and also Greece. Ancient haplotype [298-311] was also observed in the Czech Republic, Bavaria, Bulgaria, France and Syria. Haplotype [294-304-320] occurred in Germany (central and south) and both the modern and ancient population of England. In the latter, this haplotype belonged to a (female) individual of the Roman settlement of Leicester.

No.	hg	haplotye	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
1	H'	crs	7	11	25	10	8	55	32	3	13	145	39	5	70	1	21	4	29	1	2	3	3	4	28	4	33	7	19	9	44	16	14	31	18	6	5	21	21	12	10	14	13	10	171	14	56	27	26	23	9	26	3	1181	51
2	J'	126	3	3	2	1	3	7	2	2	1	26	9	2	24		2	3	5	1	1	1	1		7		1		1	3	5	1	2		11		3	1	1	4	1	1	1		66	2	17	6	9	7	4	7	3	263	44
5	H'	311	3	5	2	9	3		1	6	2	8	3	4	14	1	2	2	2	1		2		3	4		2	2	3		5	1	3	2	5	1	2	2	9	1	1		5	2	16	1	4	3	1	2		2	1	153	44
6	H'	189	2	1	3	1	1	4	3	8		7	6		13		2	2	12	1	2	1	1	1	9	2	1		1		1	2		1	8			1		1			3		16	2	5	3	3		1	2	1	134	38
8	T2	126 294 296 304	1		1	1		7	2		3	15	5	1	4		1	3	3		1		2	1	1	1					4	1	1			1				1	1			26		4		2				3	98	29
9	T1	126 163 186 189 294	1	3	2		2	2	2	3		4	3	2	2		3		1	1	1		1	9			3				2	2			1	2	2	4	1	2	1	1	4	2	11	2	3	1	2	1			2	91	37
10	H'	129		1	1	2		2				7	1		28				2						1		2	1	2		4			6	1		1			3		1			8	1	7	1			1		1	85	24
11	I	129 223		1	2			3	1	1	1	5	4		9		1			1					1	1									1		1						1		21	5			2			10	1	73	21
13	U5a1a	256 270		2		2		2		2		4	3		3		4		1		1	1									1				3	1		1		1	2		13			1	2	1	2	4	2	60	25
14	H'	162	1					3	2			6	5	1	2	3	5		7		1	1		2							1						1	3		1				5	2	2			1		3	2	60	23
15	U5a1	192 256 270		5	2	1		2		1		9	3	1	5	1			1			1			2						3				3					2			10	1	2		1				1	57	21
20	U5b	189 270					1	2	1	2		7			6	3	1		2	1		1			2	1	1				1				2				2			2		2		1		1			2	44	22
24	W	223 292	1				2	1	2			6	3						1							1			1		2					2		2						6		2	1	3				1	37	17
35	T5	126 153 294			1				1			6			12																				1															1	1	23	7
40	U5a1	192 256 270 291						1				2	2		2				2			1			1																	1			5	1	1			1		1	1	22	14
42	H'	221				2	1	1				1			1											1		2		1	1				1						1		1		1	1		3			1	20	16
45	U5a1	192 256 270 311			1							1							1				1																2		2		1	5		2					1	1	2	20	12
108	H'	222						3		1	1																														1													1	7	5
132	V	298 311		1			1																		1								1			1																	1	6	6
191	H'	294 304 320						1	1																																							1						1	4	4
220	X	183c 189 223 278																															1			1																	1	3	3
412	K	224 293 311																																												1								1	2	2
414	T4	126 187 294 296 324																																												1								1	2	2
417	B?	172 187 189 217 223																																												1								1	2	2
432	H'	162 292																																																				2	2	1
433	V	218 239a 298																									1																												1	2	2
473	W	184 223 292																	1																																				1	2	2
1079	IXW	192 223 274 301																																																				1	1	1
1080	J'	126 293c																																																			1	1	1
1081	H'	129 192 311																																																			1	1	1
1082	H'	183c 189 239																																																			1	1	1
1083	X	189a 223 271 278																																																			1	1	1
1084	T2	126 192 260 294 304																																																			1	1	1
1085	T2	126 192 207 294 304																																																			1	1	1
1086	T1	126 163 186 189 193d 294																																																			1	1	1
1087	T1	126 163 186 193d 239 294																																																			1	1	1

Table V.3 Haplotype sharing pattern for the 36 HVS-I haplotypes from ancient Britain. Haplotypes were defined by nucleotide differences with the CRS (Anderson *et al.* ,1981) Sites are transitions unless specified (e.g. 239a: adenine transversion and 193d: deletion). (No.) Number of haplotypes corresponding to the full dataset presented in Appendix F. Haplogroup classification was done according to the algorithm described in Methods. (∑) total number of individuals observed per haplotype and (pops) number of populations in which that haplotype was found. CODES – (po): Polonia, (cz): Czech Republic, (au): Austria, (sz): Switzerland, (bv): Bavaria, (gs): south Germany, (gc): central Germany, (dn): Denmark, (nw): Norway, (fn): Finland, (sw): Sweden, (ic): Iceland, (sa): Saami, (ka): Karelia, (vf): Volga-Finn region, (es): Estonia, (be): Belarus, (nr): European Russia, (ur): Uralic region of Russia, (kr): Ukraine, (ws): west Siberia, (bl): Bulgaria, (rm): Rome, (si): Sicily, (ni): north Italy, (sd): Sardinia, (tc): Tuscany, (sp): Spain, (pt): Portugal (fr): France, (bq): Basque Country, (gk): Greece, (sy): Syria, (tk): Turkey, (ar): Armenia, (ge): Georgia, (pl): Palestine, (ad): Adygei, (gy): Gypsy, (ro): Romania, (os): Ossetia, (sc): Scotland, (ok): Orkney, (en): England, (cw): Cornwall, (ir): Ireland, (wa): Wales, (ky): Is. Skye, (wi): Western Is., and (anc): ancient Britain

Haplotype [222] was shared between ancient Britain (a female from the late Saxon settlement of Norwich), Germany and Denmark and surprisingly also Palestine

Each of the remaining 5 haplotypes observed in ancient Britain was shared with only one modern population. Of these, three haplotypes, namely [224-293c-311], [126-187-294-296-324] and [172-187-189-217-223], one from the settlement of Norton and two from Norwich, only occurred in Scotland. The individual from the early Saxon site is a rich male aged 25-35 and the individuals from the late Saxon site are a young female and a child. Haplotype [184-223-292] was observed only in the early Anglo-Saxon settlement of Norton and in Estonia. An interesting case is haplotype [218-239a-298] which was only found in the Romano-British site and Sicily. The individual was a male aged >45 years, possibly a warrior of the Roman army.

As mentioned, the remaining 10 HVS-I sequences were private to the ancient populations. Their possible origin was investigated by searching the database for sequences differing by one –or the minimum possible– segregating sites. For most of the *private* sequences found in ancient Britain there were either a) too many sequences differing by one segregating site, or b) just one sequence observed in too many populations so that geographical origin could not be inferred from such data.

However, in a few cases these searches yielded informative results. The pair of private haplotypes [126-192-260 (or 207) 294-304] from the late Saxon settlement of Norwich both proved to be one mutational step away from haplotype [126-192-294-304] only observed in Norway and the Western Is. For the other pair of private haplotypes from Norwich [126-163-186-189-193d-294] and [126-163-186-193d-239-294] a one-step away haplotype, namely [126-163-186-193d-294], was observed only in Orkney. In addition, haplotype [183-189-239-298], one mutational step away from the ancient English haplotype [183-189-239], was observed only in the modern English population. Another haplotype, also one step away from this latter ancient one, was observed in Scotland.

Sequence w6 from the late Anglo-Saxon settlement of Norwich (a young child) presents a transversion T→A at position 189. This transversion is extremely rare, having only been found in the modern Gypsy population. Interestingly, this rare nucleotide variant is not a one-off case but actually rather frequent (8%) in this population. Two

Gypsy haplotypes presented the A→T transversion, namely [126-189a-223-278] and [189a-223-278] which are one and two mutational steps away from the ancient HVS-I haplotype [189a-223-271-278].

Given that similarities between the ancient and Gypsy populations were not expected more detailed investigation of the pattern of haplotype sharing of the Gypsy population was carried out. As expected, the Roma-Gypsies, who are located in the Balkan area, shared haplotypes with their source/home populations Romania and Bulgaria. They also shared haplotypes with other neighbouring populations from the Caucasus such as Adygei, north Ossetia and Georgia and from north-east Europe such as Estonia and Belarus. However, 'Gypsy' haplotypes were also observed in Scotland, Ireland and modern England.

V.2.3 Founder analyses

The analyses carried out here were based on the study by Helgason et al. (2001). The authors analysed mtDNA from the islands of the North Atlantic to identify their proportion of Celtic and Norse ancestry. For this, the mutational divergence of the North Atlantic mtDNA pools from possible source European populations was estimated by using the index ρ. The ρ index is defined as the average number of substitutions between the sequences of one population and the closest founder sequences observed in the putative source population. In this way the overlap between the mtDNA pools of the founder and the potential source populations can be quantified. In their study, Norway, Denmark and Sweden were considered the Scandinavian source populations and, on the other hand, Scotland and Ireland were considered the Celtic source populations. For comparisons, ρ distances were also calculated to other European populations believed to have less influence on the islands of the North Atlantic. Based on their results, the authors claimed that for all the populations of the North Atlantic Sea, the 'Celtic source' was the closest genetically related, followed second by the 'Scandinavian source'. However, a closer analysis of their data roused concern about the legitimacy of these results. By definition, founder analyses depend on the identification of founder haplotypes in the putative source populations. As the probability of finding such

haplotypes should be higher for larger populations, ρ distances would be expected to be affected by sample size. Helgason et al. (2001) stated that "sequences from geographically proximate populations were combined into groups to obtain *equivalent sample sizes*", yet source populations ranged in one order of magnitude (from 102 to 1028). Not surprisingly, the Celts (n = 1028) were the closest to all the founder populations. Moreover, the ρ distances to following populations seemed to be proportional to their sizes. The authors mentioned the possible effect of sample size in the calculation of ρ distances, but claimed that the relative position of the source populations should still be an indicator of closeness to the founder population.

To further investigate the effect of the sample size on the calculation of ρ distances, a trial test was carried out. Rho distances were computed between the ancient population of England (the *founder* population) and several modern populations of unequal sizes (the *source* populations). As in Helgason *et al.* (2001), Scotland, the largest sample (in this case, n = 891) showed the smallest ρ distance. As expected, large samples produced small ρ distance values, whereas the opposite was obtained for small populations. The only exceptions to this were the Saami and the Basques, which despite of being medium size populations (n = 130 and n = 105, respectively) showed rather large ρ distances. By this test, it was confirmed that ρ distance, calculated as done by Helgason *et al.* (2001), is strongly (inversely) correlated with sample size of the source population ($r = 0.70^{**}$, $P = 0.01$).

To solve the effect of the sample size on the founder analysis a heuristic approach was undertaken. It consisted in re-sampling the source populations so that they were all of equal size (n = 50, set to match the ancient sample) (see Methods II.3.4 for details). Rho distances between the ancient and several source populations thus corrected for sample size were computed. No correlation between ρ distances calculated in this manner and their original sample sizes was observed ($r = 0.04$, NS). The effect of correcting for sample size can be seen graphically in Figure V.5.

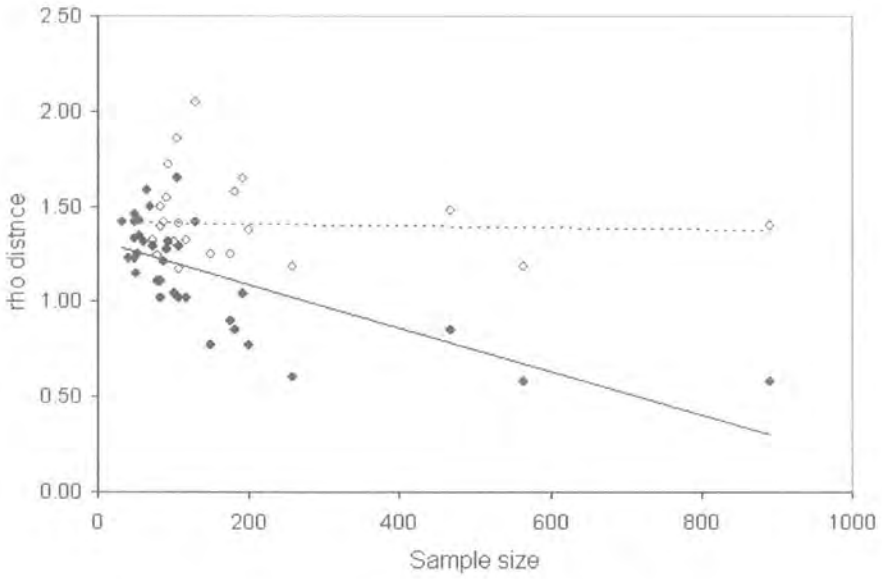


Figure V.5 Effect of sample size of the source population in the computation of ρ distances. For the same founder population (ancient England), ρ distances were calculated to the same source populations: a) corrected (clear circles) and b) non-corrected (black circles) for sample size. Trend lines are indicated for corrected (doted line, $r = 0.04$, NS) and non-corrected (full line, $r = 0.70^{**}$, $P = 0.01$).

The fist column in Table V.4 shows the ρ distances to the putative source populations (corrected for sample size) for the total ancient population of Britain, where the smallest values (at the top of the table) indicate the closest distance to the founder population. The corrected values clearly showed that the previous small ρ distance obtained for Scotland was only due to its large sample size. Instead, the closest source populations are the Caucasian population of Adygei, the modern population of England and Norway. The Gypsy population was the least genetically related to the ancient population of England, followed by the Basques. This agrees with previous results obtained for F_{ST} and Nei's distances. Interestingly, also Wales, believed to represent ancient Britons, had a large ρ distance to the ancient population of Britain as a whole.

To further investigate the origin of the ancient HVS-I haplotypes, the ancient population was divided into sub-groups. Three *founder* populations were defined according to their historical period: the *Romans,* comprising the Romano-British settlement of Leicester (4^{th} C.), the *early Saxons*, including the Anglo-Saxon settlements of Norton, Buckland, and Lavington (5^{th}-7^{th} C.) and the *late Saxons* composed by the settlement of Norwich (9^{th}-11^{th} C.). Rho distances between these founder populations and putative source populations are shown in Table V.4.

FOUNDERS							
SOURCE	**Ancient**	SOURCE	**Romans**	SOURCE	**early Sax.**	SOURCE	**late Sax.**
Adygei	1.15	England	1.00	Norway	0.86	Orkney	1.18
England	1.18	Uralic Russ.	1.00	Adygei	0.92	north Germ.	1.21
Norway	1.19	Sicily	1.17	Uralic Russ.	0.96	Palestine	1.21
north Germ.	1.20	Rome	1.17	Estonia	1.00	Portugal	1.22
Syria	1.23	France	1.17	Denmark	1.00	England	1.27
Eur.Russia	1.23	Norway	1.22	Finland	1.01	Adygei	1.29
Orkney	1.24	Estonia	1.28	Ireland	1.04	Eur.Russia	1.29
Estonia	1.25	Turkey	1.33	Syria	1.08	Syria	1.35
Finland	1.25	Eur.Russia	1.33	north Germ.	1.10	Turkey	1.41
Uralic Russ.	1.25	Ireland	1.42	Sweden	1.16	Sweden	1.41
Ireland	1.31	south Germ.	1.46	Eur. Russia	1.16	Rome	1.41
Sweden	1.31	Finland	1.50	England	1.17	Belarus	1.44
Palestine	1.32	Belarus	1.50	south Germ.	1.19	Cornwall	1.47
Turkey	1.32	Syria	1.50	Iceland	1.19	Armenia	1.49
Rome	1.33	Karelia	1.50	Scotland	1.20	Ossetia	1.50
Belarus	1.35	Spain	1.54	Orkney	1.20	Finland	1.51
south Germ.	1.38	Scotland	1.55	Czech	1.20	Is. Skye	1.59
Karelia	1.40	north Germ.	1.58	Volga-Finn	1.20	Volga-Finn	1.59
Scotland	1.40	Orkney	1.58	Belarus	1.24	Karelia	1.59
Ossetia	1.41	Ossetia	1.58	Is. Skye	1.24	Estonia	1.61
Cornwall	1.42	Palestine	1.58	Karelia	1.24	Wales	1.62
Is. Skye	1.42	Iceland	1.63	Turkey	1.26	south Germ.	1.63
Volga-Finn	1.42	Basque	1.67	Cornwall	1.28	Scotland	1.64
Portugal	1.43	Adygei	1.67	north Italy	1.28	north Italy	1.65
Iceland	1.48	Denmark	1.67	Ossetia	1.30	Norway	1.66
Czech	1.50	Is. Skye	1.67	Rome	1.32	Ireland	1.68
north Italy	1.50	Sweden	1.67	Palestine	1.34	Czech	1.75
Sicily	1.52	Portugal	1.67	Spain	1.37	Uralic Russ.	1.76
Denmark	1.56	Cornwall	1.83	Sicily	1.40	Sicily	1.82
Spain	1.57	Volga-Finn	1.83	Basque	1.50	France	1.82
France	1.58	Armenia	1.96	France	1.52	Iceland	1.85
Armenia	1.65	Czech	2.00	Portugal	1.52	Spain	1.88
Wales	1.72	north Italy	2.00	Armenia	1.69	Gypsy	2.24
Basque	1.85	Wales	2.00	Wales	1.72	Denmark	2.25
Gypsy	2.05	Gypsy	2.06	Gypsy	1.92	Basque	2.44

Table V.4 Rho distances between the ancient population of England (*founder* population) and several modern populations (*source* populations). The ancient population was divided in three groups according to their historical period: the *Romano-British* group included the settlement of Leicester (4th C.), the *early Saxon* included the settlements of Buckland, Lavington and Norton (5th-7th C.) and *late Saxon* included the settlement of Norwich (9th-11th C.). (Ancient) includes all the samples. For all source populations the re-sampling approach to correct for unequal sample sizes was undertaken, with n = 50.

The source populations showing the smallest ρ distance to the Romano-British settlement were modern England and the Uralic region of Russia. Also very closely related were Rome, Sicily and France, followed by some Nordic populations. On the other hand, the population of Wales, and to some extent the Cornish population, were not related to the Romano-British population.

Surprisingly, for the early Saxon settlements the closest source population was Norway, followed by two populations from the Caucasus region: Adygei and the Uralic Russia and several northernmost populations: Estonia, Denmark, Finland and north Germany. Modern England was less closely related, although still showed a low ρ value. Wales was also genetically distant to the early Saxon population.

The ρ distances for the late Saxon group also yielded unexpected results. The isle of Orkney was the source population which showed the closest distance. It was followed by north Germany, and strangely Palestine and Portugal. The modern population of England proved to be closely related to the late Saxon site but Norway, which showed the closest distance to the early Saxon sites, was very little associated. In any case it is worth noting that ρ distances for the late Saxon site were in general values much larger than those obtained for the early Saxon sites, indicating less close genetic relationship. Relative position and clustering of ρ distances can be visualised in Figure V.6.

In order to investigate the contribution of the ancient population to the current population of England, similar founder analyses were carried out for the population of modern England as *founder* population. Hence, for this, the ancient population of England was included as *source* population. Rho distances are shown in Table V.5 (first column). As for the early Saxon sites, the closest source population was Norway, followed by other North Atlantic and Nordic populations. The ancient population was also genetically close to the modern population of England. The (allegedly) Celtic populations, Wales and Ireland showed rather large ρ distances. Among the source populations analysed, the Basque were the least related.

Given the unexpected close relationship between Nordic populations and both ancient and modern England, a control test was carried out to corroborate that results were not due to hidden sample effects. For this, Orkney and Portugal and were analysed as *founder* populations against a number of *source* populations, including Norway.

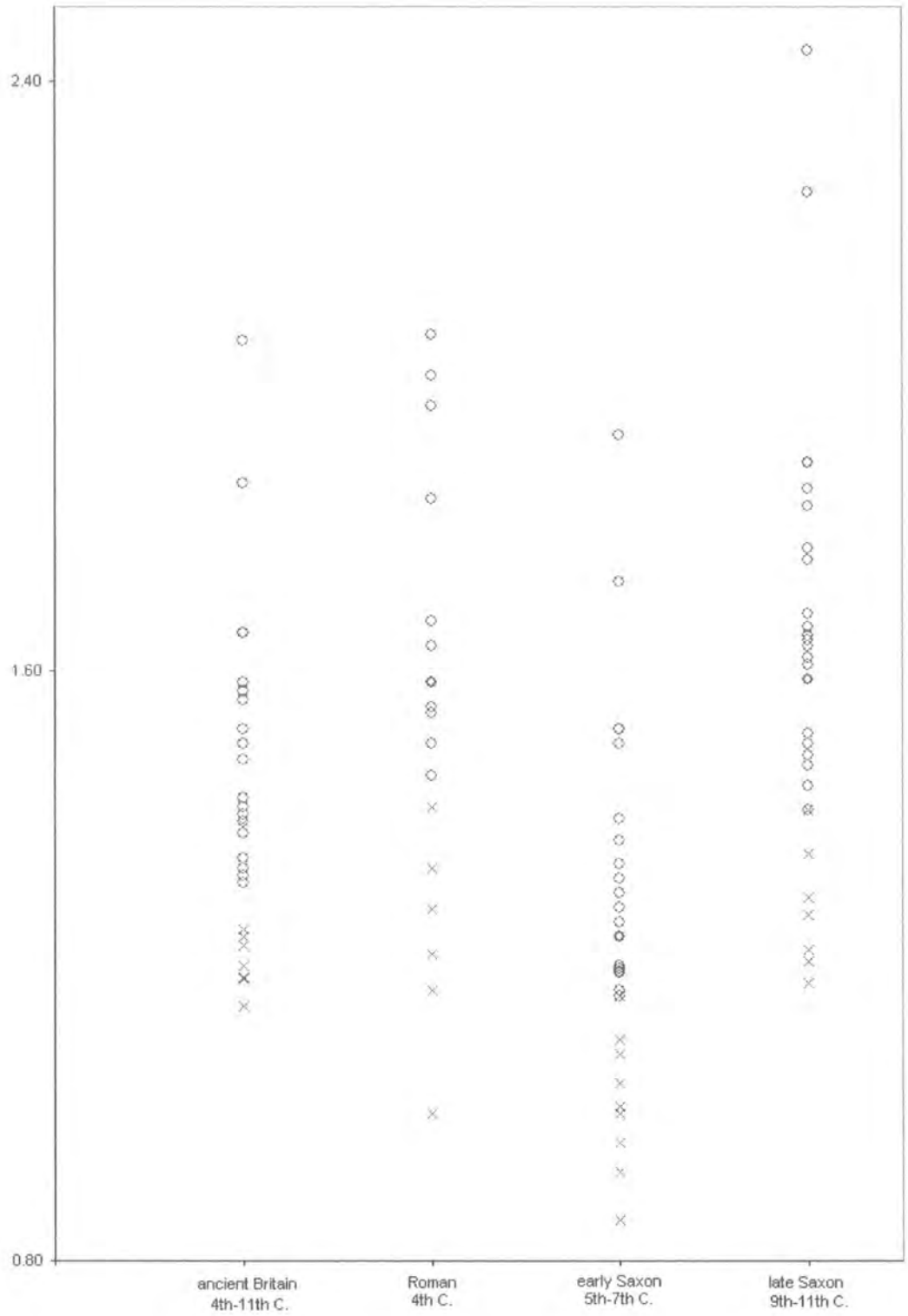


Figure V.6 Graphic representation of ρ distances and their relative position for each founder population (entire ancient sample, Romans, early Saxons and late Saxons). NOTE– Closest ρ distances are at bottom of the graph. Order of the ρ distances is shown as in Table V.4. Ten first ρ values are represented with crosses and subsequent ρ values are represented with open circles.

	FOUNDER
SOURCE	**England**
Norway	0.89
Estonia	0.94
Scotland	0.97
Finland	0.98
Orkney	0.98
Ancient Britain	**1.00**
Uralic Russia	1.00
Cornwall	1.01
Denmark	1.02
Sweden	1.02
Portugal	1.03
Syria	1.06
north Germ.	1.07
France	1.07
Wales	1.09
Turkey	1.10
Adygei	1.11
Is. Skye	1.12
Ireland	1.16
Basques	1.24

	FOUNDER
SOURCE	**Orkney**
Uralic Russia	0.72
Norway	0.85
Wales	0.92
Scotland	0.92
Palestine	0.94
Is. Skye	0.96
England	0.99
Syria	1.00
Western Is.	1.01
Finland	1.02
Turkey	1.04
Cornwall	1.06
Adygei	1.06
Estonia	1.12
France	1.14
Portugal	1.17
Ireland	1.18
Basques	1.21
Sweden	1.23
Denmark	1.29

	FOUNDER
SOURCE	**Portugal**
Adygei	0.87
Palestine	0.87
Syria	0.87
Orkney	0.89
Uralic Russia	0.89
Cornwall	0.90
Spain	0.91
France	0.92
Estonia	0.94
Sweden	0.96
Wales	0.96
Finland	0.96
Scotland	0.99
Norway	1.00
Turkey	1.01
Is. Skye	1.04
Ireland	1.06
Basques	1.10
Denmark	1.11
England	1.19

Table V.5 Rho distances for modern populations: England, Orkney and Portugal (as *founder* population) and several modern populations (as *source* populations). For all source populations the re-sampling approach to correct for unequal sample sizes was undertaken, with n = 50.

Results are shown in Table V.5 (second and third column). A totally different picture was observed. The population genetically closest to Portugal were the populations from the Near East and also Caucasus (Syria, Palestine, Adygei and the Uralic Russia). Although results are difficult to interpret, it can clearly be seen that, as opposed to the ancient and modern populations of England, for the southern population of Portugal, Norway was more genetically distant. On the other hand, France, Cornwall and Wales were comparatively more related. In addition, in comparison, the Basque population showed a smaller ρ distance to Portugal than to modern England.

In the case of Orkney, a population with a well-known Viking influence, Norway, Scotland and also Wales and the Isle Skye are among the closest genetically related. Unexpectedly, as for all previous analyses, source populations of the Caucasus and Near East seem to be closely related to the founder populations analysed.

V.2.4 Phylogenetic relationships

Given the striking differences observed between ρ distances for the different historical periods, the phylogenetic relationships of these sub-samples were separately analysed by reduced median networks (RMN). Figure V.7 shows the RMN constructed based on the HVS-I haplotypes belonging to the early Anglo-Saxon settlements. As expected, the RMN showed a star-like structure with the CRS haplotype in the centre and branches protruding from it. The majority (65%) of the HVS-I haplotypes depicted in the RMN could be assigned to one of the major European haplogroups and sub-haplogroups. In particular, haplotypes belonging to hg JT (10%), hg W (10%), hg V (5%), hg I (5%), hg U5 (including U5a1 and U5a1a) (20%), hg T (including sub-hg T2, T3 and T5) (15%) were observed. One haplotype (5%) might be assigned to cluster IXW and the remaining 30% to H' (see Chapter IV). Haplogroup X, hg K and hg J were absent, as were sub-haplogroups U5b, T1 and T4.

To further investigate the phylogenetic relationship of the mtDNA haplotypes observed in the early period, the early Saxons were analysed together with the samples from the earlier Romano-British settlement (4th C.) (the Romano-British site was not studied separately due to its small sample size). The resulting RMN based on the 31 Romano-British and early Saxon HVS-1 haplotypes is depicted in Figure V.8. The general structure of the RMN did not change, although some modifications were observed mostly due to the incorporation of new haplotypes. Haplogroup X (4%) and sub-U5b (4%) were now present and hg V accounted for 12% of the sequences. Interestingly, the presence of a Romano-British haplotype (R13) linked the central core CRS to one early Saxon HVS-I haplotype from the settlements of Buckland (B21/B33).

In addition, a reduced median network was constructed for the late Saxon group which only included the settlement of Norwich (9th-11th C.). The most striking difference resides in the predominance of hg T and the long branches corresponding to sub-hg T1 and T2 (see Fig. V.9). Absence of hg U5 motifs is also remarkable; only one U5b haplotype (which are relatively restricted to north and north-east populations) was observed.

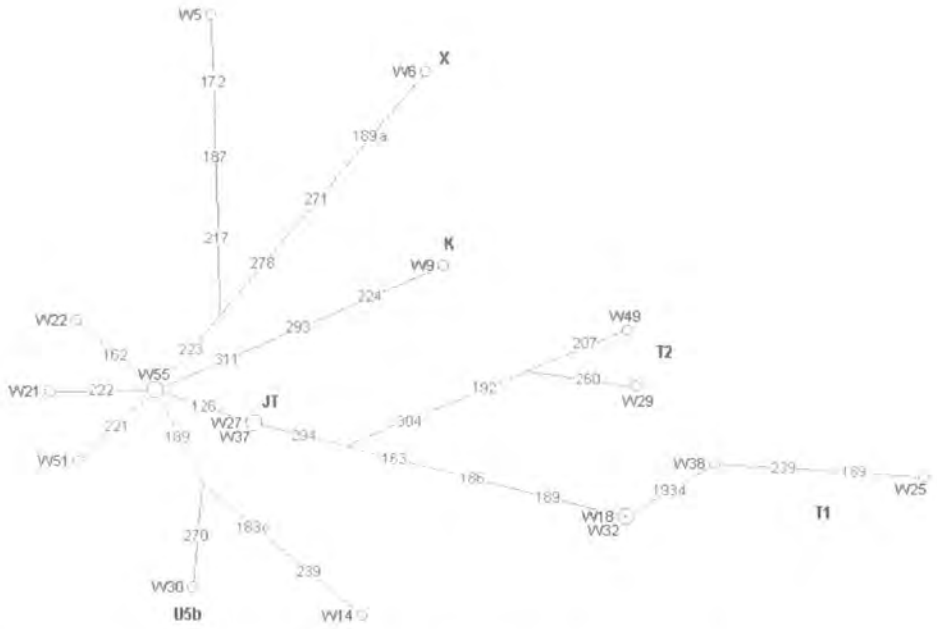


Figure V.9 Reduced median network for the *late Saxon*, the settlement of Norwich (9th-11th C.). Major haplogroups and sub-haplogroups are indicated.

Lastly, Figure V.10 shows the entire RMN, including samples from all periods as shown in Chapter IV, where the sex of the individuals is indicated. It is worth noting that no male was observed among the hg T1 and T2 samples from Norwich.

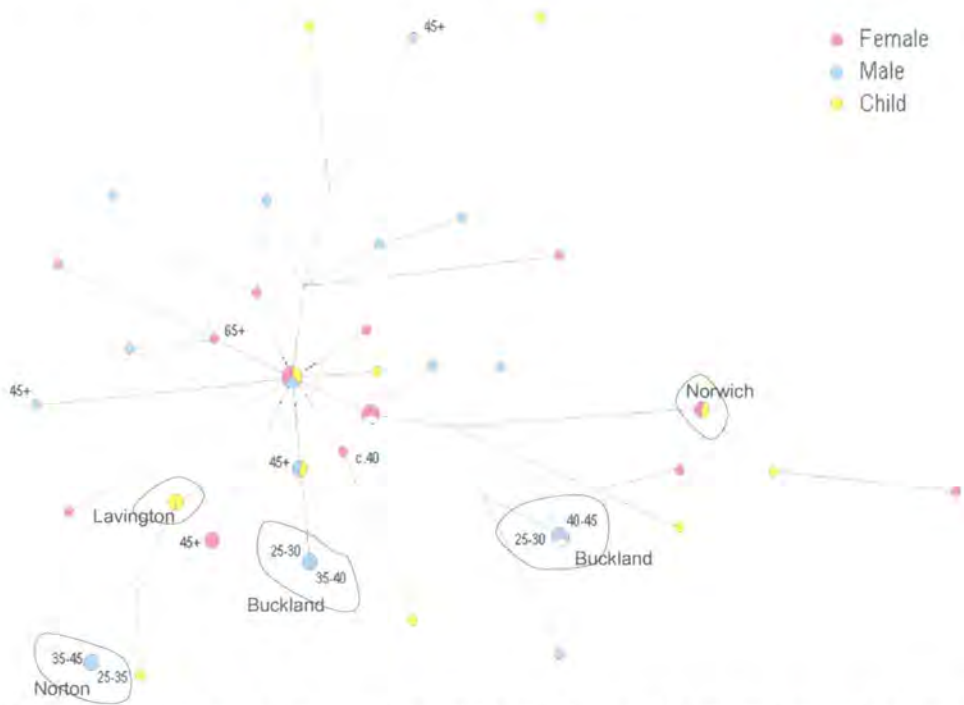


Figure V.10 Reduced median network for the ancient population of Britain, indicating the sex of the skeletal remains. Individuals of the same settlement sharing mtDNA haplotypes are further indicated.

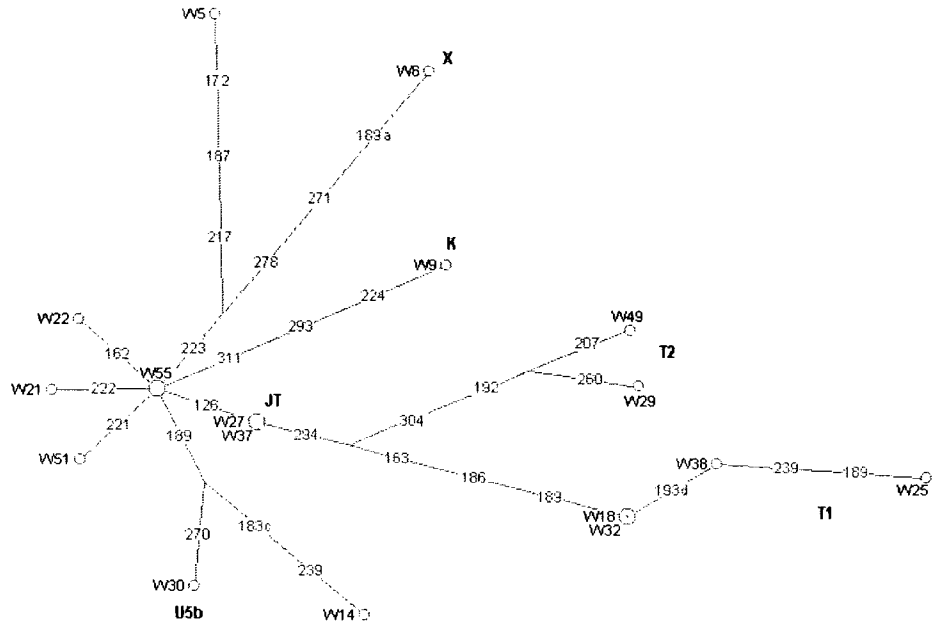


Figure V.9 Reduced median network for the *late Saxon*, the settlement of Norwich (9th-11th C.). Major haplogroups and sub-haplogroups are indicated.

Lastly, Figure V.10 shows the entire RMN, including samples from all periods as shown in Chapter IV, where the sex of the individuals is indicated. It is worth noting that no male was observed among the hg T1 and T2 samples from Norwich.

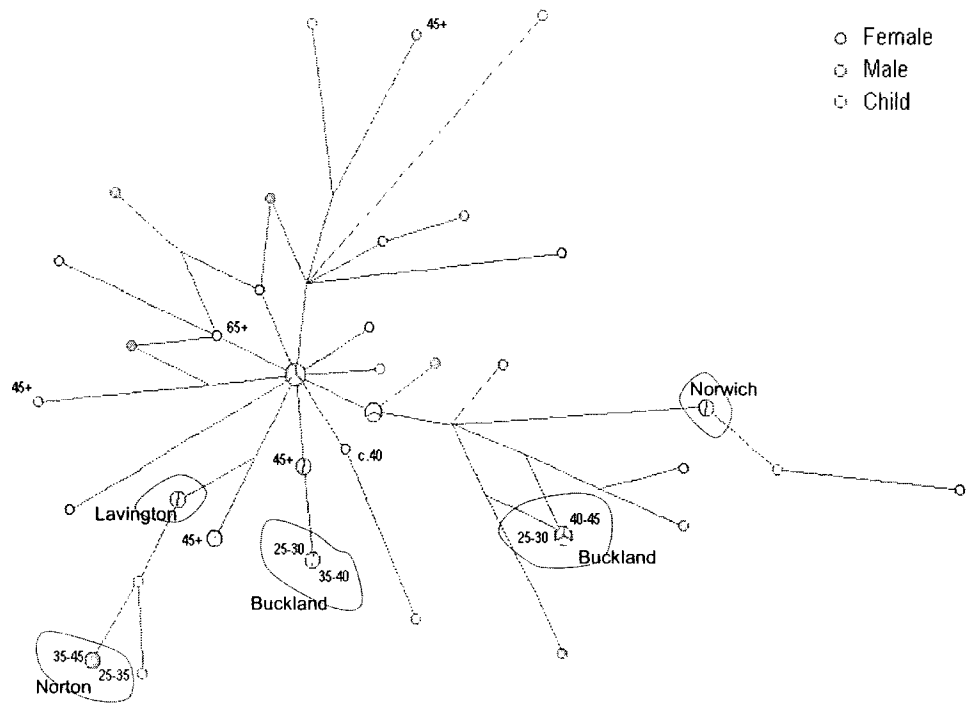


Figure V.10 Reduced median network for the ancient population of Britain, indicating the sex of the skeletal remains. Individuals of the same settlement sharing mtDNA haplotypes are further indicated.

V.3 Discussion

In this Chapter the genetic origin of populations of Anglo-Saxon England was studied. Due to the pattern of mtDNA variation in Europe, not all approaches undertaken were useful to understand the genetic relation between the ancient and modern populations. However, founder analysis, once its calculation was amended, did reveal interesting, yet unexpected results. Its interpretation will be discussed in detail.

V.3.1 Genetic differentiation

Differentiation among populations was analysed in several ways. However, neither F_{ST} statistics nor Nei's genetic distances allowed high resolution of the data as shown by the MDS plots (Figures V.1 and V.2). This had been reported in previous studies, which at the time, discouraged many researchers about the power of mtDNA analyses (Pult *et al.*, 1994; Comas *et al.*, 1996; 1997; Richards *et al.*, 1996).

Only the Saami and Gypsy populations were significantly different from all the populations ($P < 10^{-6}$), as can also be seen graphically in the MDS plots. The Basques and the Ladins of north Italy also proved to be significantly differentiated from most, although not all, other populations.

Interestingly, the ancient population of England was only significantly differentiated from the Basques (apart from the Saami and Gypsies). As a result of geographical isolation, the old proto-Basque population is thought to have had very little or no input from incoming Neolithic groups (Bertranpetit & Cavalli-Sforza, 1991; Cavalli-Sforza *et al.*, 1994). Thus, the modern Basque population has been considered to be a relict sample from old Paleolithic populations of south-western Europe (Bertranpetit *et al.*, 1995; Comas *et al.*, 1998). Based on this, Wilson et al. (2001) have argued that the male component of the Celtic populations of the British Isles (represented by Wales and Ireland) could also be seen as representing Paleolithic groups. In their work, principal component analyses of Y-chromosome haplogroup frequencies of several modern populations placed the Near Easterners and the Basques at opposite ends of the ordinate axis. These extremes were considered to represent the maximum (Near East) and minimum (Basque) contribution of early Neolithic groups to the composition of modern

populations. The Celtic populations (Wales and Ireland) plotted close to the Basques, indicating similarities between these populations. This similarity was interpreted as lack of male-mediated gene flow –through immigration– since Paleolithic times. On the other hand, when similar analyses were carried out based on mtDNA data no similarity between Basques and Celts could be detected. The authors argued then (based on the Basque~Paleolithic assumption), that the female component of the Celtic populations (Wales and Ireland) does not seem to represent the female Paleolithic inhabitants of the British Isles, and that therefore, females must have gone through more gene flow since Paleolithic times than males have. They pointed to the Neolithic and Iron Age migrations, as well as the more recent Saxon, Viking and Norman invasions as possible events responsible for such change. Accepting the Basque~Paleolithic hypotheses, the results obtained here for the F_{ST} statistics would indicate that the female component of ancient England was already differentiated from that of the Basques –and hence of the Paleolithic– female population.

As most of the paired comparisons were not statistically significant, average F_{ST} statistics and Nei's d genetic distances per population were computed. Average values for d_A showed a similar trend to those of F_{ST}, but not to d_{XY}. Difference between average d_A and d_{XY} genetic distances for a given population are due to the within-population component (d_X). High values of d_X may produce a somehow overestimated genetic distance, as measured by d_{XY}. By subtracting the within-population component, d_A estimates the genetic distance between two populations that is exclusively due to nucleotide differences which were accumulated over the time *after* population splitting –but not those already present before the split.

The population of ancient Britain had average values which ranked among the highest ($\overline{F}_{ST}$ = 0.020, $\overline{d}_{XY}$ = 5.056 and $\overline{d}_A$ = 0.097), proving to be comparatively divergent from most of the modern populations. On the other hand, modern England showed comparatively little overall differentiation ($\overline{F}_{ST}$ = 0.014, $\overline{d}_{XY}$ = 4.241 and $\overline{d}_A$ = 0.063) of which only the average $\overline{d}_{XY}$ was significantly different from the one for the ancient population ($P < 10^{-6}$). This would imply that the divergent genetic composition observed in Saxon England has faded over the past 10-15 centuries. Loss of haplotype and nucleotide diversity has already been indicated from the data in Chapter IV. It

would seem that in addition, female-mediated gene flow as a result of migration, as well as other possible processes, have homogenised, rather than differentiated, the genetic composition of England over the last millennium.

Genetic differentiation among populations was also analysed by f (chord) distances (Cavalli-Sforza & Edwards, 1967) based on haplogroup frequencies. When the distance matrix was graphically displayed by means of MDS, as for the F_{ST} and d distances graphs, the Gypsy and Saami populations fell outside the central cluster of populations. However, the haplogroup frequency analysis allowed a better resolution, as populations could be more easily singled out within the cluster. In addition, some further grouping was observed, with Nordic, south European and Near Eastern samples relatively aggregated. The ancient population of England plotted in the periphery of the cluster, in a geographically diverse area close to Denmark, Is. Skye, Cornwall and Rome among others. Interestingly, when the ancient English sample was divided in two groups according to their historical period, each group fell at opposite ends of the centre of the cluster. The early period (4^{th}-7^{th} C.) included the early Saxons and the Romano-British settlements. The late period only included the late Saxon settlement (9^{th}-11^{th} C.). Based on archaeological evidence this latter settlement was identified as Anglo-Saxon, yet Viking artifacts have also been found nearby (Penn, 2000). However, the early archaeological sites were most similar to Nordic-Scandinavian populations, including Sweden, Finland, Karelia, Denmark, Estonia, Denmark and the Volga-Finns. On the other hand, the late site plotted close to a mixture of geographic regions, including Near Eastern populations, a number of Italian populations and also the islands of Orkney, Skye and to some extent, the Western Is. Although puzzling, these results are very much in agreement to those obtained for the founder analyses and will be discussed later on.

V.2.3 Haplotype sharing pattern

Some interesting details could be seen by means of the study of haplotype sharing. First, a general analysis was carried out to identify private and shared haplotypes. As expected few haplotypes were shared between the majority of the populations and a large number of haplotypes occurred only once in the whole sample. For the private haplotypes, all populations had private *unique* haplotypes (i.e. occurring only once in the

population) and some populations also had private *multiple* haplotypes (i.e. occurring in multiple individuals from only one population). Whereas unique private haplotypes are likely due to occasional single mutations (probably along with differences in sampling effort), multiple private haplotypes could be indicative of isolation and/or differentiation. For example, the Gypsy population, known for their isolation and endogamy (Kalaydjieva *et al.*, 2001) had the largest proportion of multiple private haplotypes followed by the populations of north Ossetia, Iceland and Finland. Similar results have previously been reported for the latter two populations (Sajantila *et al.*, 1996; Kittles, 1999; Helgason *et al.*, 2000). Comparatively, the proportion of private multiple haplotypes in the Saami was not as high as in these populations. However, the isolation of the Saami was evident from the skewed frequency of common haplotypes (e.g. haplotype 4 in Appendix F). The Basques on the contrary, had a low proportion of private multiple haplotypes but a large proportion of the CRS and its derivates. Cornwall, Ireland and Wales also showed a high proportion of CRS, which was also reflected, as for the Basques, in their low average d_{XY} (Table V.2).

The ancient population of Britain had an even lower proportion of exclusive haplotypes (2.7%), suggesting that its high diversity (as seen in Chapter IV) cannot be explained by a comparatively higher proportion of private haplotypes.

Several haplotypes were shared between Nordic and populations from the British Isles and North Atlantic Islands. Some haplotypes were also shared with Caucasian populations. This agrees with results obtained for the ρ distances (see below). Among those ancient HVS-I sequences found in only one modern population, is noteworthy the case of the >45 year-old male from the Romano-British settlement whose mtDNA sequences was found only in Sicily, suggesting that the individual was a Roman soldier. Strikingly, osteological evidence is in agreement with this –the skeleton showed a fracture of the ulna well-healed and near the distal end, a lesion known as 'parry fracture' because it can be the result of placing an arm over the face to ward off a blow (Cooper, 1996).

V.3.3 Origins of mtDNA haplotypes

V.3.3.1 Methodological aspects – detecting source populations

Classical genetic distances based on nucleotide variation are computed by comparisons between all pairs of sequences from the two populations and therefore summarise the average mutational distance between them. Alternatively, rho distances are computed by identifying for each sequence of one population the *closest* putative founder sequence observed in another population (Helgason *et al.*, 2001; based on Forster *et al.*, 1996).

However, founder analyses are affected by the sample size of the putative source populations. This effect will be different depending on the type of study, mainly on whether or not the source population is known or documented. Poor sampling effort of the known source population (of a founder event) may result in an underestimation of its diversity. Hence, when compared against the founder population, most of the diversity observed in this latter would be considered new. As the accumulated diversity in the founder population is expected to be proportional to the time since its foundation, the initial underestimation of the diversity in the source population will result in an overestimation of the age of the founder event. This is of particular interest when dating age of genetic clusters believed to have originated in certain source population. (Note that this is similar for the estimation of genetic distances d_{XY} and d_A).

On the other hand, when attempting to identify the source population(s) of a colonization event, unevenness of the sample size of the putative source populations may lead to erroneous results and conclusions. As proved in this study, ρ distances are strongly correlated with sample size of the source population. Thus, for two source populations which had an equal genetic contribution into the founder population, the best-sampled population will give a smaller ρ distance, as a larger number of the founder sequences are more likely to have been sampled by chance. In the same way, ρ distances to source populations that have not been adequately sampled will tend to be overestimated. However, if poorly sampled source populations (for example, in this study, Uralic Russia with a sample size of 40) show small ρ distances, this should reflect

the actual relatedness of source and founder populations (with the only exception being back-migration and parallel mutation events).

To minimise the effect of unequal sample sizes in the calculation of ρ distances, the re-sampling method was applied in this study. This simple manual procedure allowed the identification of possible source populations which otherwise would have been obscured by their small sample size. The relative position of known source populations also proved to be altered by unequal sample sizes. Therefore, it is believed that the results obtained by Helgason et al. (2001) regarding the Scandinavian and/or Celtic contribution to the population of the North Atlantic islands were strongly influenced by differences in sample sizes. In their work, the Celtic sample was nearly twice as large as the Scandinavian one. Not surprisingly, the Celtic source presented the smallest ρ distance to all the founder populations (including Orkney, see below for details); with the Scandinavian source falling in second place.

V.3.3.2 Early colonisation

In my study founder analyses were carried out with two main purposes: first, as an attempt to identify the source populations for the ancient population of England and second, to investigate the extent to which the Anglo-Saxon population was actually the founder of the current English population.

The fact that the population of ancient England as a whole proved to be genetically close to modern England indicates that, despite what was observed in Chapter IV, at least some matrilines from that time survived to today. Interestingly, founder analyses were able to identify north Germany as a close source population to the Anglo-Saxon sample. This would confirm the Germanic nature of the settlements studied. As mentioned, two extreme models are proposed for the Anglo-Saxon invasion: total replacement or acculturation (Leeds, 1945). Based on the acculturation or integration model, Anglo-Saxon migrants and native population would have lived side by side, in an Anglo-Saxon society ruled by Anglo-Saxon government and Anglo-Saxon culture and fashion (Härke, 2001). In this scenario, Anglo-Saxon settlements, as assessed by archaeological evidence (e.g. grave finds, pottery, etc) need not be composed of Anglo-

Saxon migrants. The results obtained indicate that, at least some of the population of these Anglo-Saxon sites could have been Germanic migrants. However, nothing can be said regarding whether these were just warrior men, who immigrated in their lifetimes or whether women also crossed the North Sea with them. Rho distances computed separately for females and males might shed light on this. Unfortunately, unequal and very small sample sizes did not permit such analyses.

Although analyses of ancient England did suggest survival of some of those mtDNA lineages to present times, its relative contribution to the modern mtDNA pool was estimated to be small by means of the analyses of modern England as founder population. Gene flow by subsequent immigration events and the extinction of ancient matrilines (as seen in Chapter IV) should account for this. Given that Norway had the smallest ρ distance to the modern population of England, the Viking invasions are a strong candidate for the genetic link between the two populations. In addition, Normans may have had a strong Norse connection due to the earlier Viking settlements in Normandy (Logan, 1983). If so, post-Viking (i.e. Norman) invasion would have reinforced the mtDNA Norwegian influence –provided that women were involved in these invasions too.

Other northernmost populations such as Finland, Estonia and European Russia also showed very small ρ distances to the modern English population. Although 'Viking' connections (and influence) towards the East (and even down to Turkey) are documented, these would have been mostly trade routes rather than settlements (Noonan, 1997). Therefore further explanation seems necessary.

Clues came from further analyses of the ancient population of England. Strangely, the ancient population of Britain was found to be as close to Norway as it was to England. Given that the overall time span of ancient samples extended until 1,000 AD, the middle of the Viking period, it was believed that the small ρ distance to Norway was also due to Viking influence. To investigate this, founder analyses were carried out for sub-groups of ancient England. Three groups were defined: Romans (4^{th} C.), early Saxons (5^{th}-7^{th} C.) and late Saxons (9^{th}-11^{th} C.). The battle of Lindisfarne in 793 AD has always been considered to be the onset of the Viking Age in England. Hence, I expected that the late Saxon site of Norwich, dated 800-1000 AD was responsible for the 'Viking connection'

with the ancient sample. Surprisingly, Norway was the closest source population to the early Anglo-Saxon sites. Although some pre-Viking Age migration may have occurred they would have had to be substantial to account for such a small ρ distance. In addition, north Germany, Sweden, Finland, Estonia and European Russia also had very close ρ distances to the early Saxon sites. The widespread and northernmost geographic position of these populations suggested that the founder analyses might be highlighting a much earlier founder event.

On grounds of the lack of archaeological record during the peak of the Last Glacial Maximum, it has traditionally been believed that northern European populations migrated south in search of warmer conditions. However, it is impossible to categorically show a lack of human occupation in northern Europe. The poor archaeological record observed today might be the results of a decline in overall population, or else, environmental factors, such as heavy erosion, might have removed the existent archaeological material (Weniger, 1989). Another possibility is for any archaeological record to be currently under-water (Verhardt, 1995).

At the height of the last Ice Age, not only did the Arctic Ice cap expand to its maximum extent but also the sea levels dropped to their lowest levels. The maximum sea-level was about 125m below that of today (Fairbanks, 1989). Most of what we know of now as the North Sea was dry land, a continuation of the northern European tundra grasslands (Bjerck, 1995) and Britain was connected to the continent by a land bridge. For many years, the area of this ice-free dry land was uncertain due to controversy over the extent of the glaciation (mainly based on whether or not the Devensian and Weichselian (i.e. British and Scandinavian) ice sheets coalesced). This land was traditionally perceived as little more than a bridge between the main continent and the British 'peninsula' (Coles, 1998). However, recent advances in developments in the modelling of land-sea relations and the accumulation of data from geological surveys of the North Sea Basin indicate a more complex history of this area than imagined in the 1970's (Coles, 1999). It is now generally accepted that the area reached much farther north (Bjerck, 1995) and existed for a longer time (Coles, 1998; 1999) than had been previously thought. In addition, archaeologists have started to recognise the North Sea Continent area (also referred to as 'Doggerland') as a habitable, and more importantly,

inhabited land. In fact, archaeological evidence of traces of late-glacial human activity has and is appearing in this area (Long *et al.*, 1986; Bjerck, 1995; Thommessen, 1996; Coles, 1998). This evidence consists primarily of small occupation sites, interpreted as traces of human groups with mobile adaptation strongly dependent on large flocks of reindeer (Fischer, 1991).

In 1936, Clark produced a synthesis of archaeological evidence from Britain, north-west Europe and the North Sea Basin, suggesting that the submerged land had been the heartland of a common Mesolithic culture. Maglemosian (as he named this culture) sites which have been found from south and eastern Britain to Estonia and from Scandinavia to north Germany are not uniform, yet indicate easy contact from one end to the other (Coles, 1998). Evison (1999) also acknowledged the relative homogeneity of material culture as consistent with the notion of social relations extending across the North Sea Basin. Most Maglemosian sites represent summer and fall settlements, with small individual or nuclear family huts (Tatersall *et al.*, 1988). Clark (1936) argued that Maglemosian people migrated seasonally, since the settlements hitherto investigated seemed to be inhabited only part of the year. This scenario may mean that these people not only had a shared culture, but also shared genetic similarities.

The presence of a population living in Doggerland, does not invalidate the southern refugia model (see Introduction) as they are both compatible. It may well have been that the majority of the population of northern Europe emigrated further south (hence the scarce archaeological record), while some small communities remained in the north even when the conditions were not ideal. It has been argued that the view of northern areas totally abandoned during intense cold conditions of the LGM might ultimately be influenced by modern western concepts of comfort (Campbell, 1986) which may not reflect the capacity of adaptation that human populations possessed.

As climate and environmental conditions changed, the inhabitants of the North Sea Continent (Doggerland) may have encountered social and economic modifications. Thus, as familiar resources (e.g. reindeer) moved further north with the receding ice, humans may have followed them, or else remained and adapted to the changing conditions of the environment where they lived. Therefore, some may have moved

north-west (to England and Scotland) and some north-east, to Scandinavia and other areas of north Europe (see Fig. V.11).

Figure V.11 Hypothetical extension of the land bridge, i.e. Doggerland, during the Holocene/late Mesolithic (~10,000 YBP) and hypothetical expansion routes (modified from Jones, 1999).

Clear similarities between archaeological finds from the North Sea continent and the earliest south-west settlements of Norway (ca. 10,000 YBP), suggest that the early south settlements were populated from the North Sea Continent (Bjerck, 1995). Supporting evidence seems to be less clear for the British Isles, mainly because archaeologists have long accepted that England and Scotland had been repopulated –in that order– from the south via France, and therefore for a long time any contradictory evidence had been ignored. In any case, a number of tools and arrowheads found in Scottish sites seem to show features not only of south English/Welsh archaeological material but also of Scandinavian type (Wickam-Jones, 1994). In addition, recently, the earliest site so far discovered in Scotland (ca. 9,000 YBP) has been excavated and has presented lithic material similar to that of Scandinavian sites (Atkinson, pers.comm.). One possible explanation is that the early Scots and the Norwegians both came from Doggerland, suggesting a post-glacial connection between populations from Scandinavia and the north of Britain.

It might therefore be that the result obtained from the founder analyses of the population of ancient England is reflecting –and highlighting– a common past for early people from the British Isles, Scandinavia and other northernmost European areas. If

about 10,000 YPB these people lived together in Doggerland and surrounding areas in the north, it might be expected that the new populations formed after the split would share some genetic traits.

In support of this, the large data set compiled for this study showed that a particular sub-haplogroup, namely J1b1, is shared exclusively between the British and North Atlantic isles and Scandinavia. However, this HVS-I haplotype was not found in the ancient population of Britain. This implies that several other –although apparently less exclusive– mtDNA motifs must account for the small ρ distance between the northernmost and ancient populations. In addition, a recent genetic study of the Norwegian population proposes a common origin in Doggerland as the explanation for similarities in Y-chromosome frequencies observed between Norway and north Germany. Unfortunately, no British samples were included for comparison (Passarino *et al.*, 2002).

It might be argued that the similarities between the ancient population of England and the northern populations of Europe actually reflect Anglo-Saxon influence (based on the relationship between Norway and north Germany), however, the fact that the Romano-British sample is closely related to Norway reinforces the idea that the Scandinavian-British connection is due to an earlier link between these populations. In addition, the mismatch distribution for the ancient population of England showed an expansion date that is in agreement with an early origin (Chapter IV). In any case, the Anglo-Saxon migrants (who also might have shared a common ancestor with the Scandinavians) would be expected to contribute to the close relationship between the Nordic populations and the British Isles.

V.3.3.3 Viking influences

The founder analyses for the late Anglo-Saxon site (all samples from the settlement of Norwich) showed the smallest ρ distance to the population of modern Orkney.

Orkney is known for having a culturally mixed Pict-Norse background (Sawyer, 1997). It was the centre of the Earldom of Orkney (established by the Norwegian King in 900 AD) and the political and strategic centre of Norse activities in the North Atlantic

region (Logan, 1983). However, when the Vikings invaded, in ~800 AD, Orkney was already inhabited by the native population, the Picts. It has usually been assumed that the island was firstly explored and peopled by the Picts, who were based on the northernmost coast of Scotland (Ritchie, 1995). Yet, the population of Scotland was, at that time, a mixture of Celtic tribes: Picts, Scots, Gaels and Britons, so the native population of Orkney would also have been admixed (Owen, 1999). Given that Picts seem to have vanished from the archaeological record and that the original tongue was replaced by a Nordic language, it has been frequently assumed that the native population was totally replaced by the Vikings invaders (Ritchie, 1995). However, other interpretations argued that Vikings might have integrated in the native communities (Sawyer, 1997).

From the genetic perspective, Y-chromosome analysis has shown that the genetic composition of Orkney was intermediate between Celtic (represented by Wales and Ireland) and Norwegian samples reflecting the movement and settling of, at least, male Vikings (Wilson *et al.*, 2001). Regarding the female line, Helgason et al. (2001) claimed that the Celtic source was responsible for the majority of the Orcadian mtDNA pool. However, as mentioned above, these results are believed to be unreliable due to a lack of control for sample size. My study showed that Orcadian mtDNA lineages can be found in Norway, supporting the hypothesis of a movement of Viking women. Scotland and Wales were also very close to Orkney, which might indicate the presence of both Pict and Gaelic component in the modern Orcadian population. Provided those haplotypes are not due to Scottish and/or Welsh migration into Orkney after Viking times, this result supports the integration model where Vikings and native population lived side by side, as shown by the Y-chromosome analyses (Wilson *et al.*, 2001).

It could be argued that the close relationship between Orkney and Norway might be reflecting an earlier founder event, as hypothesised for the early settlements of England. However, in the case of Orkney, the link is exclusive to Norway (but no other northern populations), which might indicate a more recent founder event. The contributions from Scotland and Wales seem to indicate the same. However, an early (post-glacial) connection between Scandinavia and Orkney cannot be ruled out.

The strong Orcadian Viking heritage might explain its close genetic relationship to the late Saxon settlement, dated from the 9th to 11th C. By this time the Viking settlements were already established in the North Atlantic islands and in Britain (Owen, 1999). As mentioned earlier, although the site has been identified as Anglo-Saxon, Viking artifacts were also found in neighbouring sites (Penn, 2000). It may well be that the connection between late Saxon and Orkney represents a late Saxon/Viking link. The genetic composition Orkney would represent a sub-sample of the Viking invaders, similar to those that were also found in the late Saxon/Viking settlements, hence the genetic similarity between this settlement and Orkney. The fact that the late Saxon site is genetically close to Orkney but not to Norway, might be due to a founder effect and subsequent population growth and isolation in Orkney. Alternatively, the Norse invaders may have been different from other –now dominant– populations in Norway. In fact, for several species, genetic markers show two different colonisation routes for Norway (see Introduction I.1.3). Although the same pattern is not certain for humans, two major Mesolithic cultures are distinguished in Norway: the Fosna culture in the south and the Komsa culture at the northernmost point (Clark, 1975) which might indicate that the Norwegian population has a mixed origin. Originally, the south seems to have been populated from the North Sea Continent whereas the north might have been colonised via Russia (Thommessen, 1996). Y-chromosome analyses support this latter route (Zerjal *et al.*, 1997).

A number of HVS-I haplotypes found in the late Saxon site support its Viking influence. In comparison with the early Saxon settlements, the late settlement of Norwich presented a large proportion of hg T, including T1 and T2 (as seen in the RMNs, see Fig. V.9). They include two small clusters of two haplotypes, both of which have not been found so far in any other European population. Further analyses of the distribution of haplotypes in Europe indicated that the HVS-I sequences belonging to sub-hg T2 have probable founder haplotypes (i.e. one mutational step away) in Norway and the Western Is. Haplotypes belonging to sub-hg T1 have probable founder haplotypes only in Orkney. Whereas the sub-hg T2 haplotypes clearly indicate a direct link to Norway, sub-hg T1 haplotypes would indicate an indirect link. Thus, despite the fact that hg T frequency in modern Norway is very low (see Appendix E), the private

sub-hg T haplotypes found in the late Saxon site seem to have had a Norwegian or Norwegian-related origin. The overall large ρ distance between the late Saxon and Norway may be due to the remaining ancient haplotypes which are not present in the modern population of Norway. It would be expected hence, that the late Saxon site was a mixed settlement (as can also be predicted by the other source population which also showed small ρ values).

Another possible explanation for the late Saxon~Orkney~Norway link would reside in the nature of the early Viking attacks. It is believed the initial Viking activities in the British Isles involved males who might have settled and intermarried with the existing population of the North Atlantic islands. Later on, many of these admixed families took part in the large movement of the Viking era and the establishment of second settlements (Sawyer, 1997). In support of this, it has been suggested by Y-chromosome and mtDNA analyses that the Icelandic population is descendant primarly from Norse men and Celtic women (Helgason *et al.*, 2000). In Orkney, the centre of the Earldom, however, it would have been natural for whole families to move from Scandinavia –those related to the Earl and their retainers. Thus, the presence of those late Saxon mtDNA haplotypes in Orkney and Western Is. might point to movement of women *with* Vikings and not necessarily the movement of Viking women. (It is noteworthy that the above mentioned T1 and T2 haplotypes were found in two women and two children, see Fig.V.10). This might explain the distant genetic relationship between the late Saxon site and modern Norway and also the close relationship between Orkney and modern Norway.

In any case, whether or not these women were of Viking origin, it would seem that the alleged late Saxon settlement may have actually had Viking influence.

The lack of Scandinavian artifacts associated with the buried at the Norwich site should not be considered to reject the hypothesis of their Viking origin. Contrary to traditional archaeological conceptions, it has been argued that costumes and grave-goods do not necessarily indicate status or ethnic origin (Lucy, 2000). It is now proposed that what is found in a grave should be regarded as the construction of those who bury the dead, and not as reliable information about what that person was during his/her life time (Lucy, 1997).

Recent analyses of stable isotopes of Anglo-Saxon sites seem to support this view. Analyses of O- and Sr-isotopes of skeletal tissues can provide information about the person's place of origin. The link between the composition of an individual's skeletal remains and their place of origin arises from systematic variation between localities in the isotopes of different elements. Strontium variations are caused by radioactive decay over geological time and linked to the composition of rocks and minerals; oxygen differences arise due to chemical processes such as the evaporation of surface water and the precipitation of rainfall (Faure, 1986). Both elements become incorporated into skeletal tissues during formation so that their isotopic composition reflects the characteristics of the place of childhood residence. The analysis is therefore very attractive to investigate the origin of presumed first generation immigrants. Isotope analyses of an Anglo-Saxon site at West Herslerton in north Yorkshire revealed that 1/6 of the population analysed were of non-UK origins, most likely from Scandinavia (Budd *et al.*, 2002). Surprisingly, all of these corresponded to females that were distinguished by the absence of grave-finds, in marked contrast with the other females in the sample.

V.3.3.4 Other sources - Near East and Caucasus

Two other unexpected results were obtained: several founder populations had small ρ distances to Near East and to Caucasian populations. It is notable that these are regions where early migrations are believed to have originated.

The finding that source populations from the Near East, such as Syria and Palestine, proved to be so close to some of the founder European populations analysed (e.g. Orkney) might reflect the contribution of a migration from this area, probably the Neolithic wave-of-advance (Cavalli-Sforza *et al.*, 1994) (see Introduction I .1.4).

According to Richards *et al.* (2000) the principal clusters involved in this expansion process seem to have been J, T1 and T2. The frequencies of these clusters vary along different regions of Europe which, according to the authors, would be indicative of the Neolithic component of the population. A maximum contribution, determined as maximum frequency of those clusters has been found in south-east populations and a minimum contribution, as mentioned previously, in the Basque Country. The Neolithic

component in Scandinavia was also low, where the impact of the incoming farmers is thought to be relatively minor (Richards *et al.*, 2000).

Based on archaeological evidence it would seem that the spread of farming into Scandinavia and Britain was relatively delayed with respect to the rest of Europe (Dennell, 1983). The accumulation of radiocarbon dates indicated that the Neolithic started in Scandinavia and the British Isles towards 6,000 YBP, at least a millennium after its beginning in the river valleys. Understanding the British Neolithic has been controversial as there is no single area in northern Europe that can be considered to contain the cultural traditions that are ancestral to those seen in Britain. Instead, the British Neolithic seems to be a mixture of different traits that can be derived from a large area from Britanny to Jutland (Dennell, 1983). Case (1969) suggested that some of the donor cultures may have lived in coastal regions that are now submerged or eroded, i.e. Doggerland. In the same way, Coles (1999) proposed that the land bridge played a role in the delay of the onset of the Neolithic in Britain and Scandinavia. Sufficient evidence has accumulated to suggest that Doggerland remained a significant land mass for longer than previously thought, potentially into the early Neolithic (Johnson *et al.*, 1993). Coles (1999) argued that whatever the origin of the incoming Neolithic farmers, they would be expected to have no familiarity with the sea and its resources. This might have put them at a disadvantage with respect to the inhabitants of Doggerland, who in addition, might have been resistant to agricultural innovations. This in turn might have resulted in the delayed uptake of farming in Scandinavia and the British Isles.

With this respect, hg J and its derivates have been dated from ~10,000 and hence thought to represent the expansion of Neolithic farmers (Richards *et al.*, 2000). It is noteworthy that one of its sub-haplogroups, namely J1b, is exclusive to Britain and Scandinavia.

Although not as widely recognised as the Neolithic spread of farming, several authors have suggested that an expansion event from the east may have also occurred. Richards *et al.* (2000) pointed out that "*although the Caucasian data are difficult to interpret, the presence there of cluster distributions that are similar to those of Europe (and the Near East) should caution us that both Europe and the Near East could have been populated*

from a third region, perhaps closer to the extant Caucasian population or other populations in eastern Europe".

The first theory associates an expansion from this eastern area with the spread of Indo-European languages around 4,000 YBP (Gimbutas, 1970). The second explanation focuses on a much earlier event: a Late Upper Paleolithic (LUP) expansion. As mentioned in the Introduction, since the majority of the archaeological evidence comes from the western refugia, it has been believed that this refuge played the most important role in the repopulation of northern areas of Europe. However, as Soffer (1985) points out this is due to a lack of archaeological work in the east rather than lack of archaeological evidence. Also Hewitt (1999; 2000) suggested the importance of an eastern refuge in the Caucasus region in the repopulation of northern Europe for several species. This may well have been the case for man as well. Mitochondrial DNA data seem to support the expansion from a south-western refugium (Torroni *et al.*, 1998; 2001) but evidence is not as clear for an eastern refugium. So far, analysis of Y-chromosome haplogroup distribution supports a male-mediated migration from the Urals towards north-east Europe (Zerjal *et al.*, 1997, Semino *et al.*, 2000; Underhill *et al.*, 2000).

In my study, northern populations, such as Orkney and ancient England, showed the smallest ρ distances to Adygei and Uralic Russia, indicating that the majority of these founder (northern) haplotypes can be found in the Caucasus. It remains to be understood whether this might be reflecting a post-glacial recolonization of northern Europe from a Caucasian refuge or a more recent expansion of a Kurgan or Bronze Age culture.

Whichever the case, the Gypsy-like mtDNA lineage (found in the ancient population of Britain) seems to be a relict haplotype of a migration event from the Balkans/Caucasus area. The rare T→A transversion has only been found (and at a high frequency) among a Gypsy population suggesting a connection between the two populations. This group of Gypsies, currently inhabiting Bulgaria, are descendents of the early migration into the area from the east Balkans. The unusually high frequency of this haplotype and its derivates in the Gypsy population can be explained by their pronounced isolation and endogamy. Several Gypsy communities are found among the population of England, Scotland and Ireland today. However, Gypsies as such, are not

known to have arrived to the British Isles until the beginning of the 16th century (Fraser, 1992). Thus, the 'Gypsy-like' haplotype must have come from the east in an earlier migration.

The results obtained in my analysis seem to indicate that founder analyses allow the identification of founder haplotypes of both recent and older founder events. The genetic make up of the populations analysed could be traced back to different expansion processes in Europe. Two of these, namely the expansions which originated in the Caucasus and the Near East, have already been recognised and would be expected to be common to most European populations. The third component, however, namely an expansion from a northern area during late post-glacial time (probably exclusive to northernmost populations of Europe) has not been properly formulated until now.

Chapter VI - General discussion

The main aim of this research has been to contribute to our understanding of the history of the population of Britain. This has been carried out by directly exploring the mitochondrial DNA composition of human communities living in Britain during the 4th to 11th century.

The first challenge of this work was to obtain a sufficient number of aDNA sequences for a population-level study, and this was successfully achieved. As seen in Chapter III, this required optimisation of the methodology. Once a population sample size was obtained, the relevant hypothesis could be addressed by different approaches.

Mass male Anglo-Saxon migration has been claimed based on comparative analyses of Y-chromosome haplogroups from modern Welsh, English and Frisian populations (Weale *et al.*, 2002). If large numbers of females also migrated with the invaders, and displaced the native population, then matrilineal continuity between Anglo-Saxon communities and the modern English population would have been expected –provided my Anglo-Saxon sites are composed exclusively of Saxons. Ancient mtDNA types representing Anglo-Saxon communities would be expected to be well-represented among the modern population, or if extinct, closely related to extant modern types. A number of common mtDNA haplotypes are shared between the ancient and modern populations of England, yet several haplotypes (28%) are no longer present in the modern population. On the other hand, founder analyses show that the modern population of England reflects a mixture of source populations. In particular, the modern population seems to have been founded only in part by its ancient counterpart, as other populations show closer genetic distances. If the individuals in the settlements studied do represent Saxon invaders, then the female Saxon contribution to the modern English population seems to have been relatively small. On the other hand, if they represent native Britons –living in an Anglo-Saxon culture– then the contribution of those females seem to have been relatively displaced.

It was expected that the genetic diversity of the population of Britain would increase since Anglo-Saxon times, encompassing not only Anglo-Saxons but also all subsequent waves of migration to the island (as is accounted for by the founder analyses). Counter-intuitively, the opposite is observed, with the ancient population of Britain showing higher values of both haplotype and nucleotide diversity. A loss –rather than a gain– of diversity has therefore, occurred over the last millennium. As discussed in Chapter IV, loss of mtDNA diversity may have been due to variations in family size due to differential survival through epidemics in the Medieval period. This also may account for some of the lack of continuity between the ancient and modern populations.

The diversity observed in the ancient communities of Britain is much higher than most samples from modern Europe, ranking together with Near Eastern and Caucasian populations. These localities have been considered to represent places of origin of early human migrations (Cavalli-Sforza *et al.*, 1994; Richards *et al.*, 2000). Thus, the level of diversity observed in the Anglo-Saxon communities might actually represent the original diversity observed in ancient Europe which has been homogenised over the time. In support of this, mtDNA diversity observed in ancient Basque populations was also comparatively higher than the modern Basque estimate (Izagirre & de la Rúa, 1999).

All but one haplotype observed in the ancient communities of Britain can be classified into the expected major mtDNA haplogroups for European populations. Interestingly, that sequence, which might belong to the Amerindian haplogroup B, has also been found in Scotland. The ancient hg U5, believed to have originated in Europe during the Upper Paleolithic ~50,000 YBP (Richards *et al.*, 1998; 2000) was very frequent among the early Saxon communities but nearly absent among the late Saxons. On the other hand, hg T1 and T2, believed to represent most of the Neolithic expansion beginning ~10,000 in the Near East (Richards *et al.*, 2000), were more common in the late Saxon communities. Another Neolithic haplotype, sub-haplogroup J1b1, was not found in any of the ancient settlements. However, this haplogroup represents only <1.5% of modern English samples and may not be represented among the ancient samples simply due to the sample size. Nevertheless, this haplogroup proved to be Scandinavian-British Isles specific, highlighting a common past for these two localities (Chapter V and see below).

No mtDNA marker could be assigned uniqualy to Britons or invader populations. However, genetic distances measured by the rho statistic allowed a better understanding of the relationship between *founder* and *source* populations, by identifying the founder haplotypes and eliminating 'noise' due to common haplotypes. This is particularly important when analysing shallow diversity as is the case for mtDNA, where the divergence between populations is small relative to the overall mutational divergence across Europe (Helgason *et al.*, 2001). Whereas rho distances, in conjunction with detailed haplotype sharing analyses, did allow the recognition of Viking influence in the late Saxon settlement, no strong relationship between the early Saxon communities and German populations (presumably representing Saxons) was found. Instead, the early Saxon communities seem to reflect a much earlier event which might imply that these early sites represent primarily the indigenous population that existed in England prior to the Anglo-Saxon invasion. If so, the relative lack of continuity between these ancient settlements and modern England (see above) would indicate that the native population was partially replaced by subsequent invasions (especially by the Vikings) or that their genetic footprint had been erased over time.

The rho data further indicated a distant relationship between the early ancient sites and modern Wales, even though Wales has been suggested to represent the population of pre-Anglo-Saxon Britain (Hunter Blair, 1956). One possible interpretation is that differentiation between England and Wales pre-dates the Anglo-Saxon invasion. With respect to this, Weale et al. (2002), in their Y-chromosome study, warned that their analysis does not allow them to distinguish between a male Anglo-Saxon migration, from other events that added to the original genetic composition of the indigenous males of Britain. Wilson et al. (2001) have shown that the Y-chromosome haplogroup distribution of Celtic-speaking populations of Britain (Welsh and Irish) is strikingly similar to that of the Basques. Based on this and on the assumption that the Basque represent a Paleolithic relict, they argued that the male genetic impact of subsequent immigration events –both the Neolithic and the more recent invasions– must have been small. On the other hand, the female component measured as haplogroup mtDNA frequencies seemed to have undergone more gene flow since then. The more recent invasions (specially the Vikings on grounds of the results for the late Saxon site) are obvious candidates, yet an earlier event would also account for this differentiation.

It has traditionally been believed that after the depopulation during the Last Glacial Maximum (LGM) Britain was recolonized from the south, probably via France (Clark, 1932). Based on the close genetic relationship between the early Saxon populations and the northernmost populations of Europe –and supported by new approaches to archaeological evidence (Bjerck, 1995; Verhardt, 1995; Coles, 1998)–, I propose that the re-colonisation of Britain after the LGM could have been from an area now submerged under the North Sea (referred to as Doggerland by Coles, 1998). Mesolithic communities may have lived in this area, and when the climate improved, moved west and east to re-populate Britain and Scandinavia, respectively. This theory does not invalidate the traditional view that some groups migrated further south, from where subsequent expansion also occurred. In fact, Britain could have been re-populated by different groups of people through different routes.

Although speculative, it might be that the ancient British and Welsh populations represent two different recolonization routes to the British Isles. The ancient Welsh might have come from the south with a population sheltering in southern refuge (hence their Basque similarities?) whereas the ancient English might have come from the east from the area now submerged. If these two populations remained relatively isolated a pattern like that seen by Weale et al. (2002) would be expected. In addition, if it were only the men of these populations who remained isolated but the women mixed, a pattern like the one seen by Wilson et al. (2001) would be expected. These two scenarios are not mutually exclusive: they only require exclusive female migration from east to west in prehistoric times after the re-occupation of Britain occurred.

It has been proposed elsewhere that Doggerland may have existed as a considerable land for longer than previously thought, until it was flooded by the augment in the sea-level (Coles, 1999). Late Mesolithic groups and early Neolithic groups might have encountered and interacted with each other in this region. If so, this might explain the late development and slight genetic contribution of the Neolithic period in Britain and Scandinavia (Mirazón Lahr *et al.*, 2000). The shared Neolithic mtDNA haplogroup (i.e. J1b1), between these populations is in agreement with this. It is noteworthy that this haplogroup is also observed in modern Ireland and Wales, suggesting that Neolithic females did move to the west.

It has been argued that the role of males in migration has been exaggerated and that when it comes to exchanging genes among populations, women's movements have been much more important than previously thought (Stoneking, 1998). Two factors seem potentially important: one is the tendency, at marriage, for women to migrate more than men, since women more often relocate to join their spouse. In anthropological terminology, marriage is more often than not patrilocal or virilocal (Winick, 1970). This is believed to have been true throughout human history, from hunter-gatherers (Ember, 1978) to Medieval times (e.g Gerstenberger, 2002). In addition, the effect of sex-specific migration in influencing human genetic variation has been confirmed in modern populations of Thailand (Oota *et al.*, 2001). The other factor, known as exogamy, is the tendency to 'marry out' of their group, and a particular form of this, hypergamy, is the tendency to marry into an equal or better class. Hypergamy is observed in some traditional societies, such as modern hunter-gatherer communities (Cavalli-Sforza & Cavalli-Sforza, 1995) and Indian castes (Tambiah, 1973), and has been shown to result in a higher female gene flow between social groups (Bamshad *et al.*, 1998). Furthermore, exogamy has been long thought to be an homogenization factor in prehistoric cultures (Hawkes, 1967; Case, 1969).

Therefore, in the long term women seem to be more mobile than men even though their average daily displacement may be less than that of men. In agreement with this, comparisons of global Y-chromosome and mtDNA data have shown that the male markers tend to be more localized geographically than the female markers (Seielstad *et al.*, 1998). Thus, regarding the issue of the phylogeography of mtDNA, it may be that the higher female mobility over time has homogenised and blurred some previously existing geographic patterns. This together with the other demographic processes which may have altered the genetic composition of ancient populations (as discussed in Chapter IV), may result in the comparatively poor geographic mtDNA differentiation seen today. However, if appropriate methods are used, some of the underlying pattern may be distinguished. In addition, population-level analysis of ancient populations, as the one carried out in this research, allows us to have a direct appreciation of the past. Hence, further study of pre-Saxon Welsh and British sites for mtDNA and, if possible, Y-chromosome markers would help to elucidate this issue.

My results have shown that some of the mtDNA diversity present in ancient communities living in England during the 4th to 11th century has been lost over the past millennium. In addition, they showed that the early and late Anglo-Saxon sites had relatively different female components. Whereas the late Anglo-Saxon females seem to reflect Viking connections, the females from the early sites had no affinity with Germanic people. These females may therefore represent the native population of Britain, which shared a common past with other populations across the North. During the Last Glacial Maximum some of the pre-existing inhabitants of Britain may have stayed in the North in areas now submerged that would have been still suitable for settlement. When the climate improved these populations could have re-colonised northern territories. The diversity observed and data for mismatch distributions for the ancient sites are consistent with the hypothesis that they reflect an ancient expansion. Subsequent invasions and loss of diversity due to demographic events (possibly driven by epidemics), could result in the matriline composition seen in England today.

Appendix A - aDNA extraction protocols, buffers and solutions.

1 - InstaGene™ matrix (BioRad)

Dental powder was incubated with 200 µl of InstaGene™ matrix (Chelex 5% solution) (previously thoroughly stirred) for 15 min. at room temperature and for 45 min. at 56 °C, with agitation every 5 min. to avoid precipitation of the resin. Samples were vortexed for 1 min. and incubated for 8 min. at 95 °C to 100 °C. The Chelex resin was spun down by centrifugation at 10,000g for 3 min. The DNA in solution was ready for amplification. Titration of incubation times was carried out. Sets of samples were incubated for 15, 30, 60 and 90 min. at room temperature and at 56 °C.

2 - Silica I protocol (Evison et al., 1997)

Dental powder was incubated with 1 ml of 0.5 M EDTA and 25 µl Proteinase K (20 mg/ml) for 48 hr. at room temperature. Samples were pelleted by centrifugation for 5 min. Aliquots of 500 µl of supernatant, i.e. aqueous DNA solution, were transferred to clean tubes. DNA was bound to 20 µl of a silica suspension (see below) in presence of 1 ml 4 M GuSCN for 2 hr. at room temperature. The caotropic agent GuSCN was used to enable the formation of the silica-DNA compounds. These compounds were pelleted by centrifugation at high speed for 30 sec. The supernatant was discarded taking care not to disturb the pellet. The DNA-silica compounds were washed to eliminate the remaining caotropic agent. 1.5 µl 70% Ethanol was added to the tubes, pellets resuspended by vortexing, and pelleted by centrifugation at high speed for 30 sec. Supernatants were carefully discarded and the ethanol wash repeated. To eliminate all traces of GuSCN, a final wash with acetone was done. Pellets were dried out at 56 °C in a block heater for 5 min. Next, 110 µl of distilled water were added and the DNA eluted from the silica by heating at 56 °C for 15 min. and vortexed every 5 min. The samples were centrifuge at high speed for 2 min. for the free silica to precipitate. Approximately 100 µl of DNA solution were taken off with a pipette taking care not to dislodge the silica and transferred to a new clean tube.

3 - Silica II protocol (modified from Boom et al., 1990)

Tooth fragments were incubated with 1 ml of GuSCN-L6 buffer (see below) over night at room temperature. Samples were pelleted by centrifugation at high speed for 5 min. Aliquots of 500 µl of supernatants (i.e. aqueous DNA solution) were transferred to a clean tube. DNA was bound to 40 µl of a silica suspension in presence of 500 µl of fresh GuSCN-L6 buffer (containing the caotropic agent) by incubation at room temperature for 15 min. DNA-silica compounds were pelleted by centrifugation at high speed for 10 sec. The supernatant was discarded taking care not to disturb the pellet. The DNA-silica compounds were washed to eliminate the remaining caotropic agent. 500 µl of a solution of 70% ethanol, 0.2 M NaCl, 10 mM EDTA was added to the tubes; pellets resuspended by vortexing and precipitated by centrifugation at high speed for 15 sec. Supernatants were carefully discarded and washed twice with 500 µl of 70% ethanol. To eliminate all traces of the caotropic agent, pellets were washed with 500 µl acetone. Pellets were dried out at 52 °C for 10 min. 65 µl of TE buffer was added and the DNA eluted from the silica by heating at 52 °C for 15 min. with occasional vortex. The silica was pelleted by centrifugation at high speed for 2 min. Approximately 60 µl of DNA solution were taken off with a pipette taking care not to disturb the silica pellet and transferred to a new tube. Another 65 µl of TE were added to the silica and the elution process repeated. The two eluates aliquots were pooled together, yielding a final volume of DNA solution of 130 µl approximately.

4 - GeneClean for Ancient DNA protocol (Bio101 Inc.)

Tooth fragments were incubated with 1 ml of 0.5 M EDTA, 50 µl 10% SDS and 50 µl of proteinase K 20 mg/ml) for 15hs at room temperature. 500 µl of DeHybernation solution A were added and the mixture incubated at 60 °C for 4 hr. Samples were briefly centrifuge and aliquots of 500 µl of the supernatant were transferred to a new tube. 300 µl GlassMilk and 1 ml of fresh DeHybernation solution A were added and the samples incubated at 40°C for 2 hr. Samples were centrifuged at 4000g for 20 sec. and the supernatant discarded. 500 µl of Salton Wash 1 were added and the tubes mixed by vortex to resuspend the pellet. Samples were loaded into a spin filter and centrifuged. The filtered washing solution was discarded. 500 µl of Salton Wash 2 were loaded into the column to further wash the pellet retained on the filter. Spin columns were centrifuged and the second washing solution discarded. 500 µl of Alcohol Wash were added to into the column and filter through by centrifugation for 2 min. Spin filters were placed into clean catch tubes and 65 µl of Elution Solution added. The pellet was carefully resuspended by hand. Another 65 µl were added and the elution step repeated. The two eluted aliquots were pooled together, yielding a final volume of DNA solution of 130 µl approximately.

5 - Modified QIAamp tissue kit protocol (Harvey, pers.comm.)

Tooth fragments were incubated with 800 µl of MH cutting lysis buffer (see below) and 50 µl of Proteinase K (20 mg/ml) for 48 hr at room temperature. 20 µl of Proteinase K of the same concentration were added half way through the soak. Samples were briefly centrifuged and aliquots of 500 µl of the supernatant were transferred to a clean tube. 500 µl ATL buffer (QIAgen) was added and the tubes incubated at 70 °C for 10 min. 410 µl of ethanol were added and mixed thoroughly. Half of the volume of this mixture was loaded into a spin column and centrifuged at 8,000 rpm for 30 sec. The filtered solution was discarded and the rest of the mixture was loaded into the filter and centrifuged for 1 min. Collection tubes containing the flowthrough solution were discarded and spin columns placed into new tubes. 500 µl of AW buffer were loaded and samples centrifuged to filter the washing solution through. The wash was repeated two times. DNA was eluted from the filter in 100 µl of preheated water by incubating the spin columns at 70 °C for 5min. Samples were briefly centrifuged and the eluate was collected and re-heated at 70 °C to be re-used for a second elution step, yielding a final DNA solution volume of approximately 100 µl.

6 - QIAquick Spin columns protocol (Yang et al., 1998)

Dental powder was incubated overnight at 55 °C with 5 ml of lysis buffer (0.5 M EDTA pH 8.0, 0.5% SDS) and 50 µl of Proteinase K (20 mg/ml) and then at 37°C for another 24 hr. Samples were centrifuged for 5 min at 2,000 g and aliquots of 1.5 ml of supernatants were transferred to clean 10 ml tubes. 5 volumes of QIAquick PB buffer (QIAgen) were added to the tube and mixed. 750 µl of mix were loaded into a QIAquick PCR purification column (QIAgen) and centrifuged for at high speed for 45 sec. The flowthrough was discarded and the process repeated until all the extract was passed through the column. DNA was washed by adding 750 µl of QIAquick PE buffer and centrifuged for 1 min. at high speed. The flowthrough was discarded and the samples were centrifuged for 2 min. to eliminate any trace of ethanol. DNA was eluted from the columns by loading 100 µl TE buffer and centrifuging for 1 min.

7 - Phenol/chloroform and Silica-column extraction (modified from Richards & Sykes, 1995)

Tooth fragments were incubated with 1 ml 0.5 M EDTA pH 8.0 on a rotary wheel at room temperature overnight. 1% SDS was then added to a final concentration of 0.5% and 50 µl of Proteinase K (20 mg/ml) and incubated for at least 10 hr. at 37 °C. Samples were centrifuged for 5 min at 2,000 g and aliquots of 0.5 ml of supernatants were transferred to clean tubes. The DNA was extracted with an equal volume of equilibrated phenol, phenol/chloroform and chloroform/isoamyl alcohol at 24:1, following (Sambrook, *et al.* 1989), except that – due to the high EDTA concentration – during the first extraction the DNA aqueous phase was the lower phase. DNA aqueous solutions were concentrated by filtering through silica-based spin columns following Yang et al. (1998) as above.

Solutions and Buffers

1x TE buffer
10 mM Tris-HCl (pH 7.4)
1 mM Na_2EDTA (pH 8)

Size fractionated Silica (Boom *et al.*, 1990)
Dilute 4.8 g SiO_2 in 40 ml distilled water (total volume), vortex and sediment by leaving it on lab bench for 24 hr. at room temperature. Remove 35 ml of supernatant and fill up with distilled water to 40 ml, then resuspend pellet by vortexing. Sediment for further 5 hr at room temperature and remove 36 ml of the supernatant. Add 48 µl HCl (32% w/v) (approx. pH 2). Aliquot into 0.5 ml tubes.

GuSCN-L6 buffer (Boom *et al.*, 1990)
120 g GuSCN
100 ml 0.1 M Tris-HCl (pH 6.4)
22 ml 0.2 M EDTA (pH 8.0)
2.6 g Triton X-100

MH Cutting Lysis buffer (Harvey, pers.comm.)
5ml 0.5 M EDTA, pH 8
5 ml 10% SDS
250 µl 1 mM DTT
250 µl 0.5 mM Spermidine, in 50 ml distilled water

RM Lysis buffer (Montiel, pers.comm.)
0.5% SDS
0.45 M EDTA pH8
10 mM Tris-HCl

Ethidium Bromide solution
EtBr was kept as a 10 mg/ml stock solution and was protected from light.

10x TBE buffer
107.8 g Tris Base
7.44 g EDTA
55 g boric acid, in 2 litres distilled water

6x Loading buffer
0.25% bromophenol blue
0.25% xylene cyanol FF
40% sucrose, in distilled water

Appendix B - Genetic identification of human skeletal remains victims of human rights abuses

CASE STUDIES

Group Case (Fátima) — on 21st August 1976 local newspapers reported the finding of twenty-five dynamited corpses on the outskirts of Fátima, Buenos Aires, Argentina. According to witnesses, the dismembered bodies were of young persons, had their hands tied and seemed to have been previously shot in the head. Police intervention took place and the corpses were collected for legal procedures. However, the bodies were disposed of in the local cemetery as NN (i.e. no name), with their date of death being the only data recorded in the cemetery files. In 1986 the EAAF took charge of the case; the so-called Fátima mass grave was dug up and the skeletal remains exhumed and taken to the laboratory for analyses. Only one of the victims, who had a pacemaker, could be identified by the standard method. Thus, DNA analysis was the only approach which might allow the identification of the remaining bodies.

Individual cases — several skeletal remains were exhumed from illegal individual graves in local cemeteries. Seven NN cases belonging to the cemeteries of Moreno (2 male NNs), La Plata (3 female NNs) and Avellaneda (one female NN), Buenos Aires and Cnel. Villegas (one female NN), Cordoba, Argentina were also studied for identification. Based on historical data, the EAAF established strong hypotheses about their identities. However, lack of pre-mortem data did not allow positive identification of any of the remains. DNA analyses of pairs [putative relative-skeletal remain] might confirm the identity hypotheses.

Putative relatives — living relatives of 'disappeared' persons, to whom these unidentified skeletal remains might belong, were contacted. Since mitochondrial DNA would be studied, only individuals who were biologically maternally related to the victims were considered for the analysis. A total of 45 dried blood spot samples were collected, 39 living relatives potentially related to the victims in the Fátima case (three of which were analysed at a later stage) and 6 related to the 7 individual cases. To avoid cross-contamination, these samples were analysed by another investigator in a different lab.

For further details of these cases as well as for more information about the EAAF's history and *modus operandi*, see http://www.eaaf.org.ar

RESULTS

Group case

DNA from all the forensic samples was extracted, amplified and sequenced as described in the main text of this thesis. Sequences for both fragments (230 bp and 380 bp) of the mtDNA control region were aligned and compared against the CRS (Anderson *et al.*, 1981). Polymorphic positions for forensic samples and living relatives are shown in Table III.4 (main text of the thesis) and Table B.1, respectively. Population frequencies of the observed haplotypes were evaluated based on a human mitocondrial DNA worldwide database (Handt *et al.*, 1998).

Among the 25 forensic samples, 22 different mtDNA haplotypes were observed. Of these, 19 were singleton haplotypes, i.e. mtDNA variants found in only one individual. The other 3 haplotypes were observed in two skeletal remains each, namely [D1-D13], [D7-D11] and [D10-D22], suggesting that these

pairs of individuals may be maternally related to each other. However, it is not possible to determine the degree of relatedness of these individuals by mtDNA analyses alone. Samples D1, D7, D11 and D13 belong to young male bodies and samples D10 and D22 to female bodies, the former to an elder woman.

Among the 39 sequences obtained for the presumed relatives of the victims of Fátima, 37 different haplotypes were observed. Of these, 35 were singletons and 2 were multiple haplotypes. These latter were shared between two individuals each (rel8 and rel9 and rel72 and rel73, respectively). Such pairs of individuals were actually maternally related to each other and therefore expected to share the same mtDNA haplotype. In particular, individuals 8 and 9 are mother and daughter and individuals 72 and 73 are siblings.

Sample	Polymorphic HVS-I sites							Polymorphic HVS-II sites					
Presumed relatives of Derqui's remains													
rel 1	189	223	325	362				73	146	207			0/+
rel 7	235	291						93					+/+
rel 8 [d]	126	224						185	228	295			0/+
rel 9 [d]	126	224						185	228	295			0/+
rel 15	182	183	189	217	249	312	344	152	271				+/+
rel 28	224	301	311					73					2+/+
rel 34	248												2+/+
rel 35								150					2+/+
rel 45								93					0/+
rel 52	223	287	298	311	325	327		194	249d	290d	291d		0/+
rel 59	311												0/+
rel 67	291	311											+/+
rel 68	189	223	298	325	327			185	249d	290d	291d		+/+
rel 69	223	241	298	301	309								+/+
rel 70													2+/+
rel 71	290	319											0/+
rel 17													+/+
rel 18	189												0/+
rel 38	178	183	189					73					0/+
rel 42	179	243						73	239				+/+
rel 47	129c	182	183	189				73	152	217			2+/+
rel 58	261							200					0/+
rel 63	248							150					2+/+
rel 72 [e]	175	213											0/+
rel 73 [e]	175	213											0/+
rel 74								195					0/+
rel 77	218	261						152					+/+
rel 83	223	290	319										
rel 84	157	223	242	311	325								
rel 85	182	183	189	241	274			73	103	152			0/+
rel 90	298							72	131				0/+
rel 93	256	270											
rel 96	304												
rel 111	233												
rel 113								73a					0/+
rel 116	183	189	217					73	143	151			0/+
rel 92	126	300						73	185	228	295		0/+
rel 107	183	189	217					n/a					
rel 112	189	223	298	325	327			249d	290d	291d			+/+
Presumed relatives of La Plata's remains													
rel 136	223	311						73					+/+
rel 149	223	242	311	325	327								+/+
rel 150	147	183	189	217				143	195				0/+
Presumed relative of Moreno's remains													
rel 82	126	163	186	189	294			73	152	195			+/+
Presumed relative of Villegas' remains													
rel 100	126	241						150	185	189	195		0/+
Presumed relative of Avellaneda's remains													
rel 36	111	223	290	319	362			73	146	150	235	263	+/+

Table B.1 Polymorphic positions along the HVS-I and HVS-II regions (total of 610 bp) for the 45 putative relatives. Numbering according to Anderson *et al.*, 1981 (for HVS-I, prefix 16 was omitted for brevity). Mutations are all transitions unless stated. (d) deletion; (+) No. of insertions at the 309/310 cytosine tract of the HVS-II. (d) and (e) indicate individuals sharing the same mtDNA haplotype; in particular relatives 8 and 9 and mother and daughter and relatives 72 and 73 are siblings.

When all the samples analysed for the Fátima group case were compared to each other, mtDNA haplotypes belonging to seven skeletal remains matched those of living relatives, suggesting a maternal biological bond between them. In particular, the skeletal remains D4, D5, D6, D8, D16, D17 and D19, matched the relatives 70, 93, 67, 59, 8 (and also 9), 77 and 68, respectively. In all the cases the sex and estimated age of the skeletal remains agreed with those expected for the missing persons. The observed mtDNA variants proved to be either unique or rare (<1%) among the mtDNAs present in a worldwide database (Handt *et al.*, 1998). This strongly confirmed the biological relationship between such individuals, hence allowing the identification of seven victims.

In two of these cases, namely [D16-rel8] and [D8-rel 59], DNA samples were also genotyped for nuclear markers. Nine short tandem repeats (STRs) nuclear loci and the Amelogenin (sex determination) locus were amplified by PCR using the human identification Amp*Fl*STR Profiler Plus™ kit (PE Applied Biosystems, Foster City, CA), according to the manufacturer's instructions, except that the forensic samples were subjected to 60 cycles. Fluorescent PCR products were analysed by automatic DNA Sequencer and Genetic Analyzer (PE ABI Prism Systems, Foster City, CA) and genotypes determined by Genotyper 2.0 software (PE Applied Biosystems, Foster City, CA). For each [skeletal remain-putative relative] pair, genotypes were compared to each other to determine the number of matching alleles. In order to weight the DNA evidence, the matching probability (or probability of match, pM) was calculated. This is the probability of finding the same DNA profile in a randomly selected individual from the population. For this, allele frequencies for each locus were taken from a STR Internet database comprising several worldwide populations and ethnic groups (Ruitberg *et al.*, 2001). As the ethnic origin of the deceased individuals was not always known, the highest allele frequencies were used and therefore, the matching probability calculated is an overestimation.

The modern samples yielded PCR product for all 9 nuclear markers, whereas for the forensic material only the smaller fragments could be amplified, as expected due to the degradation and scarcity of the DNA template (see Table B.2). Sample D16 yielded PCR products for 7 microsatellite loci. D8, however, amplified only 3 loci. For this reason the statistical significance of the test was rather low. Despite this, results for both pair of individuals supported a first-degree relatedness. The matching probability, i.e. the probability of such DNA profiles occurring by chance were 3.4×10^{-6} for [D16-rel8] and 1.6×10^{-2} for [D8-rel58], re-confirming the identity of the remains (see Table B.2. a and b). In the former case, the woman identified (i.e. D16) was already pregnant when she was kidnapped; further osteological analyses of her skeletal remains indicated that she had not given birth.

The genetic analyses also allow raising new hypotheses for the identities of some victims. This was the case for the abovementioned skeletal remains that had identical mtDNA haplotypes among each other, namely [D7-D11], [D1-D13] and [D10-D22]. When these results were communicated to the EAAF, the historical and archaeological information was studied again. For the case of the two female victims (D10 and D22), it was found out that short before the finding of the Fátima bodies, a young woman and her mother had been abducted. It would be expected that they had been killed and disposed of together. The age (at time of death) of the skeletal remains D10 and D22 was in agreement with them being a mother/daughter pair (50 ± 8 and 30 ± 5 years, respectively). To confirm this, the forensic samples were also genotyped for STR markers and a sample from a putative relative of these women (rel92) was collected for mtDNA analyses. PCR results for the nDNA analyses for these female remains corroborated their kinship and the matching probability was 1 in 7×10^6 (see Table B.2.c). Furthermore, they were also related to their putative relative, as was indicated by mtDNA analyses (see Table III.4 Table B.1), hence determining their identity.

A)

Locus designation	Remains D16 (presumed daughter)	Relative 8 (presumed mother)	Allele frequency
D3S1358	15 , **18**	17 , **18**	0.18
vWA	**15 , 18**	**15 , 18**	0.22
FGA	19.2 , **21.2**	**21.2** , 23.2	0.06
D8S1179	**12** , 14	**12** , 15	0.29
D21S11	**31** , 32.2	30 , **31**	0.10
D18S51	**12.2** , 17	**12.2** , 18	n/a
D5S818	11 , **13**	10 , **13**	0.14
D13S317	**9.2** , 11.2	**9.2** , 12.2	n/a
D7S820	**10** , 11	9 , **10**	0.35
Amelogenin	female	female	
Matching probability			3.4×10^{-6}

B)

Locus designation	Remains D8 (presumed son)	Relative 59 (presumed mother)	Allele frequency
D3S1358	15 , **16**	**16** , 18	0.25
vWA	14 , **15**	**15** , 18	0.22
D8S1179	**12** , ?	**12** , 15	0.29
Amelogenin	male	female	
Matching probability			1.6×10^{-2}

C)

Locus designation	Remains D10 (presumed mother)	Remains D22 (presumed daugther)	Allele frequency
D3S1358	**16** , 18	**16** , 17	0.25
vWA	14 , **17**	**17** , 19	0.29
FGA	24 , **29**	**29** , ?	0.02
D21S11	**29** , 31	**29** , ?	0.22
D18S51	**11.2** , 13.2	9.2 , **11.2**	0.13
D13S317	10.2 , **11.2**	9.2 , **11.2**	n/a
THO1	**7** , ?	**7** , 8	0.25
Amelogenin	female	female	
Matching probability			7.7×10^{-6}

Table B.2 Genotypes for the nuclear markers and the amelogenein gene (sex determination) for three pairs of individuals, (**A**) skeletal D16 remains and her presumed mother, (**B**) skeletal remains D8 and his presumed mother, and (**C**) skeletal remains D10 and D22 (presumed mother and presumed daughter, respectively). Alleles inherited from the presumed mother are indicated in bold. Allele frequencies were obtained from an STR database (Ruitberg *et al.*, 2001).

Regarding the male remains, when they were more carefully examined and compared to each other, it was noteworthy that the male skeletons D7 and D11 belonged to two young males (25 ± 5 and 20 ± 2 years old) both of short stature (167 ± 2 and 164 ± 3 cm). The other pair of remains, D1 and D13, belonged to two rather tall (182 ± 3 and 179 ± 4 cm) young adult males (33 ± 5 and 31 ± 4 years) who both showed a particular feature in their bottom right molars. These physical characteristics, as well as their age-range suggested they might be close relatives, such as siblings or cousins. Historical data indicated that very few pairs of siblings had gone missing together. Strong evidence pointed to two particular cases so relatives of these presumed individuals were contacted for subsequent analyses. A sample (rel112) was obtained for the pair [D7-D11]), but unfortunately, there are no maternal relative of the pair [D1-D13] alive.

Individual cases

For the individual cases of La Plata, Moreno, Cnel. Villegas and Avellaneda cemeteries, all 7 human remains had different mtDNA haplotypes. Likewise, for the putative relatives, 6 different haplotypes were observed (see Table B.1). Mitochondrial DNA sequences of the presumed relatives and the human remains were compared for each case. In 5 out of the 7 cases, the haplotype observed in the skeletal remains perfectly matched that observed in the putative relative counterpart. Thus, the genetic analyses allowed the corroboration of the initial hypotheses of identity. In particular, the NN skeletal remains H13, G23, C26, V796 and D8-42 were identified.

CONCLUSIONS

The methods developed in the preliminary aDNA study proved to be suitable for the analyses of skeletal remains. Genetic material was successfully extracted from all the forensic samples, amplified and sequenced. Primers used for PCR left out some 20 bases at each end of the standard (full-length) HVS-I and HVS-II fragments. However, the resulting loss of information was not significant, as the omitted zones are the less variable regions of the D-loop (Lutz *et al.*, 2000). Thus, although the mtDNA fragments amplified were shorter than those analysed in other forensic studies (e.g. Sajantila & Budowle, 1991; Ginther *et al.*, 1992; Primorac *et al.*, 1996; Bender *et al.*, 2000; Budowle *et al.*, 2002) they showed to be adequate for the analyses.

This genetic analysis allowed the elucidation of forensic cases that could have not been resolved otherwise. Of the 31 skeletal material analysed in this study, 14 were identified. Thus, the skeletal remains D4, D5, D6, D8, D10, D16, D17, D19, D22, H13, G23, C26, V796 and D8-42 are no longer NN but José Daniel Bronzel, Juan Carlos Vera, Susana Ocampo, Horacio Garcia Castelau, Hayde Rosa Cirullo, Susana Elena Pedrini, Norma Frontini, Jorge Daniel Argente, Carmén Maria Carnaghi, Elena Arce, Nora Formiga, Adriana Delgado, Maria Leonor Papaterra and Rosario Victoria Ramirez. They have all been restored to their families, who could finally bury them and thus close a seemingly never-ending wake. With respect to this, it is believed that that the main aim of this side-project has been accomplished. Further work should be undertaken as there are still many unidentified bodies and families that need testing. The identification of victims of human rights abuses, despite its non-academic nature, is a creditable application of the aDNA methodology, and my contribution to a more just society.

Appendix C - Details of DNA extraction, PCR amplification and sequencing for the archaeological material analysed.

The following tables show the details for the DNA extraction, PCR amplification and sequencing for the archaeological material.
Indiv.: label of the sample analysed, equivalences are given in Methods.
Tooth: dental sample analysed.
Extract: powder material for DNA extraction; indicates sample, tooth number and extract number.
PCR 1/2/3: PCR result obtained for successive amplifications.
Chelex: PCR result obtained after further purification with chelating resin (see Methods).
Seq.: quality of the sequence obtained.
Polymorphic sites: nucleotide positions which differ from the CRS (Anderson *et al.*, 1981).

NOTES–

+/++/+++	quality of sequences arbitrarily rated with respect to the 'background noise'
(backg)	refers to 'background noise'
(het)	heteroplasmy was observed for the position between brackets
(blurred)	'blurred' sequence due to slippage during amplification as a result of a polycytosine tract
unread.	unreadable sequence
cont.-a	contamination with AT's DNA sequence
cont.-b	contamination with TG's DNA sequence
cont.-c	contamination with RF's DNA sequence
cont.-d	contamination with ARH's DNA sequence
cont.?	possible contamination, unknown source
artifact	aDNA artifact
rec PCR	reconstructive PCR
lab 1	sequence replicated in independent lab 1, Ancient Biomolecules Centre, Department of Zoology and Biological Anthropology, University of Oxford, UK.
lab 2	sequence replicated in independent lab 2, Smithsonian National Museum of Natural History, Washington DC, USA
not conf.	possible aDNA sequence that could not be replicated (not included in the analyses)

Authenticated aDNA sequences are shown in **bold type**.

Indiv.	tooth	extract	PCR 1	PCR 2	PCR 3	Chelex	Seq	polymorphic sites	comments
R1	1	R1.1.1	---	---	---	---			
		R1.1.2	+				++	(CRS)	cont.?
		R1.1.3	---	---	---	---			
	2	R1.2.1	---	---	---	---			
		R1.2.2	+				++	126, 294, 296, 304	
R2	1	R2.1.1	---	---	---	---			
		R2.1.2	---	+			+++	**294, 304, 320**	authenticated
		R2.1.3	---	---	---	+	unread.		
		R2.1.4	---	---	---	---			
	2	R2.2.1	+				+++	294, 304, 320	
		R2.2.2	+				+++	(145, 176, 223)	cont.-a
R3	1	R3.1.1	---	+			+++	**183c, 189, 223, 278**	authenticated
		R3.1.2	---	---	---				
	2	R3.2.1	---	---	+		+ (blurred)	183c, 189, 223, 278	reseq.
		R3.2.2	+	---			++	(126, 261, 274)	cont.-c
		R3.2.3	---	---	---	---			
	3	R3.3.1	---	---	---	---			
		R3.3.2	+				++ (het)	183c, (189), 223, 278	
R4	1	R4.1.1	+				+++	(224, 311, 320)	cont.-d
		R4.1.2	+				+++	(287)	cont.
	2	R4.2.1	---	---					
		R4.2.2	+				+++	*129c, 183, 189, ?*	
	3	R4.3.1	+				unread.		
		R4.3.2	---	---	---				
R5	1	R5.1.1	+					(287)	cont.
		R5.1.2	---	---	---	---			
	2	R5.2.1	---	---	---				
		R5.2.2	+					*126, 290, 292, 311*	?
	3	R5.3.1	---	---	---				
		R5.3.2	---	---	---				
		R5.3.3	+				++	(CRS)	??
R6	1	R6.1.1	---	---		---			
		R6.1.2	+				++	(145, 176, 223)	cont.-a
	2	R6.2.1	+				unread.		
		R6.2.2	---	---	---				
	3	R6.3.1	+				unread.		
		R6.3.2	+				++	(CRS)	cont.?
R7	1	R7.1.1	+				+++	**189, 270**	authenticated
		R7.1.2	---	---		---			
		R7.1.3	+				+ (backg)	(189) 270 (295)	
	2	R7.2.1	+				+++	189, 270	
		R7.2.2	---						
		R7.2.3	---						
R8	1	R8.1.1	---	---	---	---			
		R8.1.2	+				+++	**126, 294, 296, 304**	authenticated
		R8.1.3	---			+	unread.		
	2	R8.2.1	+				+++	126, 294, 296, 304	
		R8.2.2	---	---					
R9	1	R9.1.1	---	---		---			
	2	R9.2.1	+				++ (het)	218, 239a, (298)	
		R9.2.2	---	---	+		+++	239a, 298	shorter fragment
	3	R9.3.1	+				+++	**218, 239a, 298**	
		R9.3.2	+				++	(CRS)	cont.

(continued)

Table C.1 Extraction, PCR and sequencing details for the dental samples from Newarke Street cemetery at Leicester. Codes are as described on the first page of this Appendix.

Indiv.	tooth	extract	PCR 1	PCR 2	PCR 3	Chelex	Seq	polymorphic sites	comments
R10	1	R10.1.1	---	---		+	unread.		
		R10.1.2	---	---	---	---			
		R10.1.3	---	---	---				
	2	R10.2.1	---	---	---	---			
		R10.2.2	+					(175, 213)	cont.?
		R10.2.3	+				+ (het)	(175) 183g (213) 290	??
R11	1	R11.1.1	---	---	+	---	unread.		
		R11.1.2	---	---	---				
	2	R11.2.1	---	---	---				
		R11.2.2	+				+ (het)	(126, 290)	
		R11.2.3	---	---					
R12	1	R12.1.1	---	---		---			
		R12.1.2	---						
	2	R12.2.1	---						
		R12.2.2	+				unread.		
R13	1	R13.1.1	+				+++	**162**	authenticated
		R13.1.2	---						
	2	R13.2.1	---	---		---			
		R13.2.2	+				+++	162	
R14	1	R14.1.1	---	---	---				
		R14.1.2	---	---	---				
	2	R14.2.1	+				unread.		
		R14.2.2	---	---	---				
R15	1	R15.1.1	---	---	---	---			
		R15.1.2	+				+++	(126, 261, 274)	cont.-c
		R15.1.3	---						
	2	R15.2.1	+				+ (het)	(126, 186, 191?)	artifact?
		R15.2.2	---						
	3	R15.3.1	---	---	---				
		R15.3.2	+				+ (het)	(126, 163, 186, 192?)	artifact?
		R15.3.3	+				+ (het)	(126, 186)	

Indiv.	tooth	extract	PCR 1	PCR 2	PCR 3	Chelex	Seq	polymorphic sites	comments
N1	1	N1.1.1	+				+++	**126, 187, 294, 296, 324**	authenticated
		N1.1.2	+				unread.		
		N1.1.3	---	---					
	2	N1.2.1	+				+++	126, 187, 294, 296, 324	
		N1.2.2	+				unread.		
		N1.2.3	+				+++	(145, 176, 223)	cont.-a
N2	1	N2.1.1	+				unread.		
		N2.1.2	+					(145, 176, 223)	cont.-a
		N2.1.3	---	---	---				
		N2.1.4	---	---	---				
	2	N2.2.1	+				unread.		
		N2.2.2	---	---	---				
		N2.2.3	---	---	---				
		N2.2.4	---	---	---				
N3	1	N3.1.1	+				+++	**298, 311**	authenticated
		N3.1.2	+				++ (het)	(298) 311	
	2	N3.2.1	+				+++	298, 311	
		N3.2.2	+				unread.		
		N3.2.3	---						
N4	1	N4.1.1	+					*126, 187, 294, 296, 324*	not conf.
		N4.1.2	+					(145, 176, 223)	cont.-a
		N4.1.3	+				unread.		
	2	N4.2.1	---	---	---				
		N4.2.2	+				unread.		
		N4.2.3	---	---					
N5	1	N5.1.1	---	---	---	---			
		N5.1.2	+				+++	129, 192, 311	
		N5.1.3	+				++ (het)	(129) 192, 311	
	2	N5.2.1	+				+++	**129, 192, 311**	authenticated
		N5.2.2	---						also lab 2.
		N5.2.3	---						
N6	1	N6.1.1	---	---	---				
		N6.1.2	+				unread.		
		N6.1.3	---						
	2	N6.2.1	+					*221, 291*	not conf.
		N6.2.2	---	---	---	---			
		N6.2.3	---	---	---	---			
N7	1	N7.1.1	+					**192, 256, 270, 311**	authenticated
		N7.1.2	+				+ (het)	(126) 192, 256, 270 (278) 311	
	2	N7.2.1	---	---					
		N7.2.2	---	---					
		N7.2.3	+				++ (backg)	192, 256, 270, 311	
N8	1	N8.1.1	+				+++	(CRS)	cont.?
		N8.1.2	---	+				*218, 232, 245*	not conf.
		N8.1.3	---	---		---			
	2	N8.2.1	+				unread.		
		N8.2.2	+					(145, 176, 223)	cont.-a
		N8.2.3	---	---	---				
		N8.2.4	---	---	---				
N9	1	N9.1.1	---	+			+++	**CRS**	authenticated
		N9.1.2	---	+			++ (backg)	CRS	
	2	N9.2.1	+				+++	CRS	
		N9.2.2	---						
		N9.2.3	---	+			++ (back)	CRS	

(continued)

Table C.2 Extraction, PCR and sequencing details for the dental samples from the Norton cemetery at Cleveland. Codes are as described on the first page of this Appendix.

Indiv.	tooth	extract	PCR 1	PCR 2	PCR 3	Chelex	Seq	polymorphic sites	comments
N10	1	N10.1.1	+					(145, 176, 223)	cont.-a
		N10.1.2	---						
	2	N10.2.1	---	---	---				
		N10.2.2	---	---		---			
		N10.2.3	+				unread.		
N11	1	N11.1.1	+					(248)	cont.
		N11.1.2	---						
		N11.1.3	+				+++	**192, 256, 270, 291**	authenticated
	2	N11.2.1	---	---					
		N11.2.2	+				+++	192, 256, 270, 291	
		N11.2.3	---						
		N11.2.4	---						
N12	1	N12.1.1	+					**126, 293c**	authenticated
		N12.1.2	---						
	2	N12.2.1	---	+				126, 293c	
		N12.2.2	---						
	3	N12.3.1	+					(248)	cont.
		N12.3.2	+					(175, 213)	cont.
N13	1	N13.1.1	+				+++	(126, 261, 274)	cont.-c
		N13.1.2	---						
		N13.1.3	+					(248)	
	2	N13.2.1	---						
		N13.2.2	+				(het)	126, 148, 184, 186, 223, 290, 301	artefact?
		N13.2.3					(het)	126, 201, 234, 259, 267	artefact?
		N13.2.4	---						
N14	1	N14.1.1	---	---		+	unread.		
		N14.1.2	+					*139, 256, 270*	not conf.
		N14.1.3	---	---	---	---			
		N14.1.4	---	---	---	---			
	2	N14.2.1	+					311	?
		N14.2.2	+				+++	(CRS)	cont.?
		N14.2.3	---	---	---				
N15	1	N15.1.1	+				+++	**126, 153, 294**	authenticated
		N15.1.2	+				+++	(145, 176, 223)	cont.-a
	2	N15.2.1	+				+++	126, 153, 294	
		N15.2.2	---	+			+ (het)	(126) 153, 294	
		N15.2.3	---						lab 2.
N16	1	N16.1.1	+					**129a, 223**	authenticated
		N16.1.2	---	---					
	2	N16.2.1	+					129a, 223	
		N16.2.2	---	---					
		N16.2.3	+					(126, 261, 274)	cont.-c
N17	1	N17.1.1	+				++	CRS	cont.?
		N17.1.2	+				++	126, 294, 296, 304	??
		N17.1.3	+				+++	(146, 176, 223)	cont.-a
	2	N17.2.1	---	---	---				
		N17.2.2	---	---	---				
		N17.2.3	---	---	---				
N18	1	N18.1.1	---	---					
		N18.1.2	+					(145, 176, 223)	cont.-a
		N18.1.3	---	+				(CRS)	?
	2	N18.2.1	---	---	---				
		N18.2.2	+				+ (het)	126 (294, 296)	
		N18.2.3	---	---	---				
		N18.2.4	+				++	(CRS)	?

(continued)

Indiv.	tooth	extract	PCR 1	PCR 2	PCR 3	Chelex	Seq	polymorphic sites	comments
N19	1	N19.1.1	+					**192, 256, 270, 311**	authenticated
		N19.1.2	+				+ (het)	192, 256, 270, (290/2), 311	
		N19.1.3	---						
	2	N19.2.1	+					192, 256, 270, 311	
		N19.2.2	---						
		N19.2.3	---						
		N19.2.4	---	+			+ (het)	(184/5) 192, 256, 270, 311	
N20	1	N20.1.1	+					**184, 223, 292**	authenticated
		N20.1.2	---	+			+ (het)	223, 292 (294)	
	2	N20.2.1	+					184, 223, 292	
		N20.2.2	---						
N57	1	N57.1.1	+				+++	**223, 292**	authenticated
		N57.1.2	---	---					
	2	N57.2.1	+				+++	223, 292	
		N57.2.2	---	---					
N120	1	N120.1.1	+				unread.		
		N120.1.2	+				unread.		
	2	N120.2.1	---	---	---				
		N120.2.2	---	---	---				
		N120.2.3	+				+++	(175, 213)	cont.?
N41	1	N41.1.1	+				+++	(145, 176, 223)	rec PCR
		N41.1.2	---	---	---				
	2	N41.2.1	+				unread.		

Indiv.	tooth	extract	PCR 1	PCR 2	PCR 3	Chelex	Seq	polymorphic sites	comments
B1	1	B1.1.1	---	---	---	---			
		B1.1.2	+				+++	(145, 176, 223)	cont.-a
	2	B1.2.1	+				unread.		
B2	1	B2.1.1	+				+++	*126, 163, 186, 189, 294*	not conf.
		B2.1.2	---	---	---				
		B2.1.3	---	---	---				
	2	B2.2.1	---	---	+		unread.		
		B2.2.2	---	---	---	---			
B3	1	B3.1.1	+				+++	*126, 261*	not conf.
		B3.1.2	---						
	2	B3.2.1	+				unread.		
		B3.2.2	+				unread.		
		B3.2.3	---	---	---	---			
B4	1	B4.1.1	+				+++	*167, 189*	not conf.
		B4.1.2	---	---		+	unread.		
	2	B4.2.1	+						
B5	1	B5.1.1	+				unread.		
		B5.1.2	+				+++	(145, 176, 223)	cont.-a
	2	B5.2.1	+				+++	*126, 172, 294, 304*	?
B6	1	B6.1.1	---	+			+++	*126, 145, 294, 296, 304, 310*	not conf.
		B6.1.2	+				unread.		
	2	B6.2.1	---	---		---			
		B6.2.2	---	---					
B7	1	B7.1.1	---	---	---	---			
		B7.1.2	---	---	---	---			
	2	B7.2.1	---	---	---	---			
		B7.2.2	---	---	---	---			
B8	1	B8.1.1	---	---	---	---			
		B8.1.2	---	---	---	---			
	2	B8.2.1	---	---	---	---			
		B8.2.2	---	---	---	---			
B9	1	B9.1.1	+				+++	**126, 294, 296, 304**	authenticated
		B9.1.2	---	---	---	---			
	2	B9.2.1	---	---	---				
		B9.2.2	---	+			+++	126, 294, 296, 304	
B10	1	B10.1.1	+				+++	**126, 294, 296, 304**	authenticated
		B10.1.2	---	+			++ (het)	126 (134) 294, 296, 304	
	2	B10.2.1	+				+++	126, 294, 296, 304	
B11	1	B11.1.1	---	+			+ (backg)	**192, 223, 274, 301**	authenticated
		B11.1.2	---	---	---	+	+ (backg)	192, 223, 274, 301	
	2	B11.2.1	---	---	---	---			
		B11.2.2	+				unread.		
B12	1	B12.1.1	---	---	---	---			
		B12.1.2	---	---	---	---			
	2	B12.2.1	---	---	---	---			
B13	1	B13.1.1	+					*176, 219, 266*	not conf.
		B13.1.2	---	---					
	2	B13.2.1	---	---	---	---			
B14	1	B14.1.1	---	+			+++	**189**	authenticated
		B14.1.2	---	+			+(blurred)	189, ?	
	2	B14.2.1	+				+++	189	
B15	1	B15.1.1	---						
		B15.1.2	+				unread.		
	2	B15.2.1	+				+++	(145, 176, 223)	cont.-a
B16	1	B16.1.1	+				++	CRS	
	2	B16.2.1	---	---	---	---			

(continued)

Table C.3 Extraction, PCR and sequencing details for the dental samples from the Buckland cemetery at Kent. Codes are as described on the first page of this Appendix.

Indiv.	tooth	extract	PCR 1	PCR 2	PCR 3	Chelex	Seq	polymorphic sites	comments
B17	1	B17.1.1	---	---	---	---			
		B17.1.2	+				+++	**311**	authenticated
	2	B17.2.1	+				+++	311	
B18	1	B18.1.1	---	---	---	---			
		B18.1.2	---	---	---	---			
	2	B18.2.1	+				+++	224, 311	not conf.
B19	1	B19.1.1	+					(304?)	
	2	B19.2.1	---	---	---	---			
B20	1	B20.1.1	+					126, 294, 296, 304	
		B20.1.2	---	---	---				
	2	B20.2.1	+					(126)	cont.-b
B21	1	B21.1.1	---	---	---	---			
	2	B21.2.1	---	---	---	---			
B22	1	B22.1.1	+					**162, 292**	authenticated
	2	B22.2.1	---	---	---				
		B22.2.2	+					162, 292	
B23	1	B23.1.1	+				unread.		
	2	B23.2.1	---	---	---				
		B23.2.2	---	---	---				
B24	1	B24.1.1	+				++ (het)	(262) 298	
		B24.1.2	---	---	---	---			
	2	B24.2.1	---	---	---	---			
B25	1	B25.1.1	+				unread.		
		B25.1.2	+				+++	(145, 176, 223)	cont.-a
	2	B25.2.1	---	---					
		B25.2.2	+					298	
B26	1	B26.1.1	+				unread.		
		B26.1.2	+				+++	(145, 176, 223)	cont.-a
	2	B26.2.1	---	---	---	---			
B27	1	B27.1.1	---	---	---	---			
		B27.1.2	+				+++	(145, 176, 223)	cont.-a
	2	B27.2.1	---	---	---	---			
B28	1	B28.1.1	+				+++	(126)	cont.-b
		B28.1.2	---	---	---	---			
	2	B28.2.1	+				++ (het)	176, 186 (219)	
		B28.2.1	---	---	---	---			
B29	1	B29.1.1	+				++ (het)	(126), 168, 223, 295	
		B29.1.2	---	---	---				
	2	B29.2.1	---	---		---			
B30	1	B30.1.1	---	---	---				
		B30.1.2	---	---	---	---			
	2	B30.2.1	---	---	---	---			
		B30.2.2	+				++	(CRS)	cont.?
B31	1	B31.1.1	+				unread.		
		B31.1.2	---	+			+++	**162, 292**	authenticated
		B31.1.3	---	---	---				
	2	B31.2.1	+				+++	162, 292	

Indiv.	tooth	extract	PCR 1	PCR 2	PCR 3	Chelex	Seq	polymorphic sites	comments
L1	1	L1.1.1	---	---	---	---			
		L1.1.2	---	---	---	---			
	2	L1.2.1	+				+++	*223, 290, 319*	not conf.
		L1.2.1	---	---	---	---			
		L1.2.3	---	---	---	---			
L2	1	L2.1.1	+				+++	**CRS**	authenticated
		L2.1.2	---	+			+++	CRS	
	2	L2.2.1	+				+++	CRS	
L3	1	L3.1.1	+				+++	**192, 256, 270**	authenticated
		L3.1.2	---	---	---				
	2	L3.2.1	---	+			+++	192, 256, 270	
		L3.2.2	---						
L4	1	L4.1.1	---	---	---	---			
	2	L4.2.1	---	---	---	---			
L5	1	L5.1.1	+				unread.		
		L5.1.2	---	---	---				
	2	L5.2.1	+				+++	*311*	not conf.
		L5.2.2	---	---	---				
L6	1	L6.1.1	+				+++	(126)	cont.-b
		L6.1.2	+				+++	(126)	cont.-b
	2	L6.2.1	+				+++	311	?
		L6.2.2	---	---	---				
L7	1	L7.1.1	+				+++	**129a**	authenticated
		L7.1.2	---	+			+ (het)	129a (244)	
	2	L7.2.1	+				+++	129a	
L8	1	L8.1.1	+				++	(126)	cont.-b
	2	L8.1.2	+				+++	(145, 176, 223)	cont.-a
L9	1	L9.1.1	+				++	298	
		L9.1.2	---	---	---				
	2	L9.2.1	---	---	---	---			
L10	1	L10.1.1	+				+++	**256, 270**	authenticated
		L10.1.2	---	---	---	+	unread.		
	2	L10.2.1	---	+			+++	256, 270	
L11	1	L11.1.1	+				+++	**256, 270**	
		L11.1.2	---	+			++ (het)	128, 218, 228, 244/8, 256, 270, 278, 292, 303	artifact
	2	L11.2.1	+				+++	256, 270	
L12	1	L12.1.1	---	---	---	---			
	2	L12.2.1	+				+ (het)	(169, 319)	?
		L12.2.2	+				++	(175, 213)	cont.?
L13	1	L13.1.1	---	---	+		unread.		
	2	L13.2.1	+				+++	183/4, 192/3, 248, 259	artefact
		L13.2.2	---	---	---	---			
L14	1	L14.1.1	---	---	---	---			
	2	L14.2.1	---	---	---	---			
		L14.2.2	---	---	---	---			
L15	1	L15.1.1	+				++	(126)	cont.-b
		L15.1.2	+				+++	(146, 176, 223)	cont.-a
	2	L15.2.1	---	---	+		unread.		
L16	1	L16.1.1	+				+++	167, 223, 290, 292	artifact
	2	L16.1.2	+				+++	193, 223, 292	artifact
L17	1	L17.1.1	+				+++	126, 187, 189, 223, 264, 270, 278, 293, 301, 311	artefact
		L17.1.2	+				++	(146, 176, 223)	cont.-a
	2	L17.2.1	---	+			unread.		
		L17.2.2	+				++	(146, 176, 223)	cont.-a

(continued)

Table C.4 Extraction, PCR and sequencing details for the dental samples from the Grove Farm cemetery, Market Lavington at Salisbury. Codes are as described on the first page of this Appendix.

Indiv.	tooth	extract	PCR 1	PCR 2	PCR 3	Chelex	Seq	polymorphic sites	comments
L18	1	L18.1.1	---	---	---	---			
		L18.1.2	---	---	---	---			
	2	L18.2.1	---	---	---	---			
		L18.2.2	---	---	---	---			
L19	1	L19.1.1	+				+++	126, 172, 211, 222, 261, 290, 313	artefact
		L19.1.2	+				++	(146, 176, 223)	cont.-a
	2	L19.2.1	---	---	---			126	?
		L19.2.2	+				unread.		
L20	1	L20.1.1	---	+			++	**126**	authenticated
		L20.1.2	---	+			+++	126	
	2	L20.2.1	+				++	126	
L21	1	L21.1.1	---	---	---				
		L21.1.2	+				++	(146, 176, 223)	cont.-a
	2	L21.2.1	+				++	304	
L22	1	L22.1.1	---	+			+++	*145, 176a, 224, 311*	not conf.
		L22.1.2	---	+			unread.		
	2	L22.2.1	+				++	(126)	cont.-b
L23	1	L23.1.1	---	---	---	---			
		L23.1.2	---	---	---	---			
	2	L23.2.1	---	---	---	---			
		L23.2.2	---	---	---	---			
L24	1	L24.1.1	---	---	---	---			
		L24.1.2	+				+++	(146, 176, 223)	cont.-a
	2	L24.2.1	---	---	---	---			
		L24.2.2	---	---	---	---			
L25	1	L25.1.1	---	---		---			
		L25.1.2	---	---		---			
	2	L25.2.1	---	---		---			
L26	1	L26.1.1	---	---		---			
		L26.1.2	+				++	(146, 176, 223)	cont.-a
	2	L26.2.1	---	---	---	---			
		L26.2.2	---	---	---	+	unread.		
L27	1	L27.1.1	---	---		---			
	2	L27.2.1	+				unread.		
L28	1	L28.1.1	---	+			+++	311	?
		L28.1.2	---	---	---				
	2	L28.2.1	+				++ (het)	(126) 224, 311	?
		L28.2.2	---	---	---	---			

Indiv.	tooth	extract	PCR 1	PCR 2	PCR3	Chelex	Seq	polymorphic sites	comments
Nw1	1	w1.1.1	+				+++	270, 311	
		w1.1.2	---	---	---	+	unread.		
		w1.1.3	---	---	---	---			
		w1.1.4	---	---	---				
	2	w1.2.1	---	---	---	+	+++	128, 142, 147, 223, 232, 261, 264-8, 270, 311, 327	artifact
		w1.2.2	---	---	---	---			
		w1.2.3	+				unread.		
		w1.2.4	---	---	---	---			
Nw2	1	w2.1.1	---	---	---	---			
		w2.1.2	+				+++	(126)	cont.-b
	2	w2.1.3	---	+			---	---	
		w2.1.4	+				+++	126, 172, 267, 294, 299, 304	artifact?
Nw3	1	w2.2.1	---						
		w2.2.2	+					126, 150, 166, 216, 256, 292, 298	artifact
		w2.2.3	+				+++	(CRS)	cont.?
	2	w2.2.4	+				+++	(145, 176, 223)	cont.-a
		w2.2.5	---	---	---				
		w2.2.6	+				+++	(287)	cont.
Nw4	1	w4.1.1	---	+			+++	(CRS)	cont.?
		w4.1.2	---	---	---				
		w4.1.3	---	---	---				
	2	w4.2.1	+				+++	239	?
		w4.2.2	+				+++	(145, 176, 223)	cont.-a
		w4.2.3	+				++	(126)	cont.-b
Nw5	1	w5.1.1	+				+++	**172, 187, 189, 217, 223**	authenticated
		w5.1.2	---						
	2	w5.2.1	---	---		+	+++	172, 187, 189, 217, 223	
		w5.2.2	---						
Nw6	1	w6.1.1	---	+			+++	**189a, 223, 271, 278**	authenticated
		w6.1.2	---	---					
		w6.1.3	---	---					
	2	w6.2.1	+				+ (backg)	189a, 223, 271, 278	
		w6.2.2	+				(het)	189a, 223 (271) 278	
		w6.2.3	+				+++	189a, 223, 271, 278	
Nw7	1	w7.1.1	---	+			++	(175, 213)	cont.
		w7.1.2	---	---	---				
		w7.1.3	---	---		---			
	2	w7.2.1	---						
		w7.2.2	+				+	+	
		w7.2.3	---	+			++	(126)	cont.-b
Nw8	1	w8.1.1	---	---	---	---	---		
		w8.1.2	---	---	---	---	---		
		w8.1.3	---	---	---	---	---		
	2	w8.2.1	---	---	---	+	unread.		
		w8.2.2	---	+				(175, 213)	cont.
		w8.2.3	---	---					
		w8.2.4	+				++	(CRS)	cont.?
Nw9	1	w9.1.1	+				+++	**224, 293, 311**	authenticated
		w9.1.2	---	---	+		+ (het)	(185, 221) 224, 293, 311	artifact
		w9.1.3	+				++	(145, 176, 223)	cont.
	2	w9.2.1	+				+++	224, 293, 311	
		w9.2.2	---	---					
		w9.2.3	---	---					
		w9.2.4	---	+			+ (backg)	224, 293, 311	

(continued)

Table C.5 Extraction, PCR and sequencing details for the dental samples from the Farmer's Avenue, Castle Mall cemetery at Norwich. Codes are as described on the first page of this Appendix.

Indiv.	tooth	extract	PCR 1	PCR 2	PCR3	Chelex	Seq	polymorphic sites	comments
Nw10	1	w10.1.1	---	---	---	---			
		w10.1.2	---	---	---	---			
	2	w10.2.1	---	---	---	---			
		w10.2.2	---	---	---	---			
		w10.2.3	---	---	---	---			
Nw11	1	w11.1.1	---	+			unread.		
		w11.1.2	---	---					
		w11.1.3	---						
	2	w11.2.1	---	---					
		w11.2.2	+				++	(175, 213)	cont.
		w11.2.3	---	---					
Nw12	1	w12.1.1	---	---	---	---			
		w12.1.2	---	---	---				
		w12.1.3	+				+++	(145, 176, 223)	cont.-a
		w12.1.4	---	---	---				
	2	w12.2.1	---	---	---	---	+++		
		w12.2.2	---	+			+++	(126)	cont.-b
		w12.2.3	---	---					
		w12.2.4	---	---					
Nw13	1	w13.1.1	---	---	---				
		w13.1.2	---	---	---	---			
		w13.1.3	---	+			+ (backg)	*175, 222*	not conf.
		w13.1.4	---	---					
	2	w13.2.1	---	---	---				
		w13.2.2	---	---	---				
		w13.2.3	+				unread.		
Nw14	1	w14.1.1	+				+++	**183c, 189, 239**	authenticated
		w14.1.2	---						
		w14.1.3	---			+	++ (het)	183c, (189), 239	
		w14.1.4	+				++	(145, 176, 223)	cont.-a
	2	w14.2.1	+				+ (blurred	183c, 189, 239	reseq.
		w14.2.2	---			---			
		w14.2.3	---			---			
		w14.2.4	---						
Nw15	1	w15.1.1	+				unread.		
		w15.1.2	---	---	---				
		w15.1.3	---	---		---			
	2	w15.2.1	+				+++	(145, 176, 223)	cont.-a
		w15.2.2	---	---	---	---			
Nw16	1	w16.1.1	+				+++	*172, 223, 227, 278*	not conf.
		w16.1.2	---	---		+	unread.		
		w16.1.3	+					(126)	cont.-b
	2	w16.2.1	---						
		w16.2.2	---	+			unread.		
Nw17	1	w17.1.1	+				++	(126)	cont.-b
		w17.1.2	---	---					
		w17.1.3	---	---					
		w17.1.4	+				+++	(145, 176, 223)	cont.-a
	2	w17.2.1	---	---	---	---			
		w17.2.2	---	---	---	---			
		w17.2.3	---	--	---	---			
Nw18	1	w18.1.1	+				++	**126, 163, 186, 189, 294**	authenticated
		w18.1.2	+					(145, 176, 223)	cont.-a
		w18.1.3	---						
		w18.1.4	---	---		+	unread.		
	2	w18.2.1	+				+++	126, 163, 186, 189, 294	
		w18.2.2	---	+			+++	126, 163, 186, 189, 261/2, 270, 294	artifact
		w18.2.3	---	---	---				
		w18.2.4	+					(126)	cont.-b

(continued)

Indiv.	tooth	extract	PCR 1	PCR 2	PCR3	Chelex	Seq	polymorphic sites	comments
Nw19	1	w19.1.1	---	---	---				
		w19.1.2	---	---		---			
	2	w19.2.1	---	+			unread.		
		w19.2.2	+					(145, 176, 223)	cont.-a
		w19.2.3	+					(CRS)	cont.
Nw20	1	w20.1.1	---	---	---	---			
		w20.1.2	---	---	---				
		w20.1.3	---	---	---				
		w20.1.4	---	---		---			
		w20.1.5	+				+++	(224, 311, 320)	cont.-d
	2	w20.2.1	+				+++	*180g, 192, 256, 270, 320*	not conf.
		w20.2.2	---	---		---			
		w20.2.3	+				unread.		
		w20.2.4	---	---					
Nw21	1	w21.1.1	+				+++	**222**	authenticated
		w21.1.2	+				+++	(145, 176, 223)	cont.-a
		w21.1.3	---	---		---			
		w21.1.4	---	---					
	2	w21.2.1	+				+++	222	
		w21.2.2	---						
		w21.2.3	---						
Nw22	1	w22.1.1	---						
		w22.1.2	+				+++	**162**	authenticated
	2	w22.2.1	---						
		w22.2.2	+				+++	162	
Nw23	1	w23.1.1	---	+			+ (het)	(126, 290)	?
		w23.1.2	---						
	2	w23.2.1	---	+			unread.		
		w23.2.2	---	---	---	---			
Nw24	1	w24.1.1	---	---		---			
		w24.1.2	---	---					
	2	w24.2.1	---	---	---	---			
		w24.2.2	---	---					
Nw25	1	w25.1.1	+				++	**126, 163, 186, 193d, 239, 294**	authenticated
		w25.1.2	---	+			unread.		lab 2.
		w25.1.3	---	---					
		w25.1.4	+				++	(145, 176, 223)	cont.-a
		w25.1.5	---	+			+ (blurred	126, 163, 186, 193d, +?	
	2	w25.2.1	+				+++	126, 163, 186, 193d, 239, 294	
		w25.2.2	---	---					
		w25.2.3	---						
		w25.2.4	+				+++	(145, 176, 223)	cont.-a
Nw26	1	w26.1.1	+				unread.		
		w26.1.2	---	---	---				
		w26.1.3	---	---	---				
		w26.1.4	---	---	---				
	2	w26.2.1	+				unread.		
		w26.2.2	---	---	---				
		w26.2.3	---	---	---				
		w26.2.4	+				+++	(287)	cont.
Nw27	1	w27.1.1	---	---	---	+	++	**126**	authenticated
		w27.1.2	+				+++	126	
	2	w27.2.1	+				+++	126	
		w27.2.2	---				+++	126	
Nw28	1	w28.1.1	---	---	---	---			
		w28.1.2	---	---	---	---			
		w28.1.3	---	---	---				
	2	w28.2.1	+				unread.		
		w28.2.2	+				+++	(CRS)	

(continued)

Indiv.	tooth	extract	PCR 1	PCR 2	PCR3	Chelex	Seq	polymorphic sites	comments
Nw29	1	w29.1.1	---	+				**126, 192, 260, 294, 304**	authenticated
		w29.1.2	+				++ (het)	126, 192, (223) 260, 294, 304	
	2	w29.2.1	+				+++	126, 192, 260, 294, 304	
		w29.2.2	---	---	---				
Nw30	1	w30.1.1	---						
		w30.1.2	+				++ (het)	189 (270)	
		w30.1.3	+				++	**189, 270**	authenticated
	2	w30.2.1	---						
		w30.2.2	+				+++	189, 270	
		w30.2.3	---	+			++ (back)	(189) 270	
Nw31	1	w31.1.1	---	---	---				
		w31.1.2	+				unread.		
		w31.1.3	---	+			++	*126, 145, 231, 261*	not conf.
		w31.1.4	+				++	(145, 176, 223)	cont.-a
	2	w31.2.1	---	---	---				
		w31.2.2	---	---		---			
		w31.2.3	---	---					
Nw32	1	w32.1.1	+				+++	**126, 163, 186, 189, 294**	authenticated
		w32.1.2	---						
		w32.1.3	+					(145, 176, 223)	cont.-a
	2	w32.2.1	---						
		w32.2.2	+				++ (het)	126, 163, 186, (189), 294	reseq.
Nw33	1	w33.1.1	+				unread.		
		w33.1.2	+				unread.		
	2	w33.2.1	---	---					
		w33.2.2	---			---			
Nw34	1	w34.1.1	+				unread.		
		w34.1.2	+				++	(145, 176, 223)	cont.-a
	2	w34.2.1	---	---	---	---			
Nw35	1	w35.1.1	---	---	---				
		w35.1.2	---	---	---				
		w35.1.3	---	---	---		++		
	2	w35.2.1	+				unread.		
		w35.2.2	+				++	(126)	cont.-b
		w35.2.3	---						
Nw36	1	w36.1.1	---	---		+	unread.		
		w36.1.2	+				++	(145, 176, 223)	cont.-a
		w36.1.3	---	---					
		w36.1.4	+				unread.		
	2	w36.2.1	---	---	---	---			
		w36.2.2	---	---	---				
		w36.2.3	---	---	---				
Nw37	1	w37.1.1	+				++ (het)	126 (294)	authenticated
		w37.1.2	+				+++	**126**	
		w37.1.3	+				++	(145, 176, 223)	cont.-a
	2	w37.2.1	+				+++	126	
		w37.2.2	---	+			+++	126	
		w37.2.3	---						
Nw38	1	w38.1.1	+				+++	**126, 163, 186, 189, 193d, 294**	authenticated
	2	w38.2.1	+				+++	126, 163, 186, 189, 193d, 294	
Nw39	1	w39.1.1	---	---	---				
		w39.1.2	---	+			++ (het)	(187, 189, ?)	
		w39.1.3	---	---	---				
	2	w39.2.1	---						
		w39.2.2	---	+			unread.		
Nw40	1	w40.1.1	---	---	---	---			
		w40.1.2	+				+++	(145, 176, 223)	cont.-a
	2	w40.2.1	+				+++	(145, 176, 223)	cont.-a
		w40.2.2	---	---	---				

(continued)

Indiv.	tooth	extract	PCR 1	PCR 2	PCR3	Chelex	Seq	polymorphic sites	comments
Nw41	1	w41.1.1	+					*129, 145, 223*	not conf.
		w41.1.2	+				++	(175, 213)	cont.
	2	w41.2.1	+				unread.		
		w41.2.2	---	---	---	---			
Nw42	1	w42.1.1	---	---	---				
		w42.1.2	+				++	(175, 213)	cont.
	2	w42.2.1	---	---		---			
Nw43	1	w43.1.1	---	---					
		w43.1.2	+				++	(126)	cont.-b
	2	w43.2.1	---	---	+		unread.		
		w43.2.2	---						
Nw44	1	w44.1.1	+				++	(175, 213)	cont.
		w44.1.2	+				++	175, 190, 213	cont.?
	2	w44.2.1	---			---			
		w44.2.2	+				unread.		
Nw45	1	w45.1.1	---	---	---	---			
		w45.1.2	---	---	---				
	2	w45.2.1	---	---	---	---			
		w45.2.2	---	---	---				
Nw46	1	w46.1.1	---	---	---				
		w46.1.2	+				unread.		
	2	w46.2.1	+				+++	*130, 223, 284, 292, 325*	not conf.
Nw47	1	w47.1.1	+				+++	(145, 176, 223)	cont.-a
		w47.1.2	---	---	---				
		w47.1.3	+				++	(126)	cont.-b
	2	w47.2.1	+				unread.		
		w47.2.2	---	---	---				
		w47.2.3	---	---	---				
Nw48	1	w48.1.1	---	---	---				
		w48.1.2	---	---	---	---			
	2	w48.2.1	---	---	---				
Nw49	1	w49.1.1	+				+++	**126, 192, 207, 294, 304**	authenticated
		w49.1.2	---	---		+	+ (backg)		
	2	w49.2.2	+				+++	126, 192, 207, 294, 304	lab 1.
Nw50	1	w50.1.1	---	---	---	---			
		w50.1.2	+				++	(CRS)	?
	2	w50.2.1	+				unread.		
		w50.2.1	---			---			
Nw51	1	w51.1.1	+				+++	**221**	authenticated
	2	w52.2.1	+				+++	221	
Nw52	1	w52.1.1	---						
		w52.1.2	+				++	126, 192, 260, 294, 304	not conf.
		w52.1.3	+				+++	(CRS)	cont.
		w52.1.4	---						
	2	w52.2.1	---	+			++	126, (192), 260, 294	artifact?
		w52.2.2	---	---					
		w52.2.3	---	---					
		w52.2.4	---	---					
Nw53	1	w53.1.1	+				+++	(126)	cont.-b
		w53.1.2	---	---					
	2	w53.2.1	---	+			unread.		
Nw54	1	w54.1.1	+						
		w54.1.2	---	---	---				
	2	w54.2.1	+				+++	(126)	cont.-b
		w54.2.2	---	---		----			
		w54.2.3	---	---					
Nw55	1	w55.1.1	---	+			+++	**CRS**	authenticated
		w55.1.2	+				++	CRS	
	2	w55.2.1	+				+++	CRS	

(continued)

Indiv.	tooth	extract	PCR 1	PCR 2	PCR3	Chelex	Seq	polymorphic sites	comments
Nw56	1	w56.1.1	---	---		---			
		w56.1.2	---	---					
	2	w56.2.1	+				+++	(145, 176, 223)	cont.-a?
		w56.2.2	+				+++	(145, 176, 223)	cont.-a?
Nw57	1	57.1.1	---	---		---			
		57.1.2	---						
	2	57.2.1	+				+++	*179*	not conf.
Nw58	1	58.1.1	---	---	---	---			
		58.1.2	---	---	---				
		58.1.3	+				+++	(145, 176, 223)	cont.-a
	2	58.2.1	+				unread.		
		58.2.2	---	---	---				
Nw59	1	59.1.1	---	---		---			
		59.1.2	+				+++	(145, 176, 223)	cont.-a
		59.1.3	---						
	2	59.2.1	---			---			
		59.2.2	---			---			
		59.2.3	+				+++	(126)	cont.-b
		59.2.4	---	+			unread.		

Appendix D - Inter-population comparisons (extra tables)

Codes for the tables are as follows (in respective order)

Basq.: Basque Country
Eng. England
Corn.: Cornwall
Wal.: Wales
Scot.: Scotland
Wis.: Western Islands.
Skye: Is. Skye
Ork.: Orkeny
Denm.: Denmark
Finl.: Finland
Fran.: France
SGe.: south Germany
NGe.: north Germany
CGe.: central Germany
Bav.: Bavaria
Icel.: Iceland
Nital.: north Italy
Sard.: Sardinia
Tusc.: Tuscany
Norw.: Norway
Port.: Portugal
Saa.: Saami
Syria.: id.
Spain: id.
Swed.: Sweden
Swi.: Switzerland
Turk.: Turkey
Irel.: Ireland
Anc.: ancient Britain
Bulg.: Bulgaria
Arm.: Armenia
Bel.: Belarus
Cze.: Czech Republic
Est.: Estonia
Geo.: Georgia
Gre.: Greece
Rome: id.
Sic.: Sicily
Kar.: Karelia
Palest.: Palestine
Pol.: Poland
Rom.: Romania
Gypsy: id.
Ady.: Adygei
Oss.: Ossetia
WSib.: west Siberia
EuRus.: European Russia
Ural.: Uralic region of Russia
Ukr.: Ukraine
VFinn.: Volga-Finn region

	Aust	Basq.	Eng.	Corn.	Wal.	Scot.	Wis.	Skye	Ork.	Den.	Finl.	Fran.	SGe.	NGe.	Cger.	Bav.	Icel.
Basq.	1.16																
Eng.	0.02	0.98															
Corn.	-0.16	1.41	-0.26														
Wal.	0.27	0.99	0.10	0.20													
Scot.	0.30	1.16	0.03	-0.08	-0.04												
Wis.	0.58	**2.01**	0.59	0.45	0.85	0.75											
Skye	0.88	**4.63**	1.24	0.67	2.60	1.44	0.86										
Ork.	0.87	1.77	0.76	0.79	1.15	1.02	0.69	1.34									
Den.	1.50	**3.70**	1.13	1.21	2.09	1.34	1.78	1.41	2.61								
Finl.	**1.39**	**1.62**	**1.07**	1.19	1.97	**1.62**	**2.01**	2.64	0.56	**2.75**							
Fran.	-0.09	1.28	0.08	0.53	0.10	0.35	1.00	2.29	0.82	2.74	1.25						
SGe.	0.31	0.82	0.02	0.00	0.29	0.35	**0.99**	1.98	1.03	1.74	0.75	0.24					
NGe.	0.94	**3.35**	**1.65**	1.26	**2.70**	**2.04**	**2.17**	0.12	1.76	2.61	**2.23**	1.70	**2.06**				
Cger.	0.63	1.64	0.20	0.03	1.03	0.73	1.32	1.77	0.96	1.53	0.54	0.28	-0.01	**1.90**			
Bav.	-0.16	2.07	-0.32	-0.60	0.45	-0.18	1.04	0.50	0.29	1.10	0.66	0.13	0.22	0.16	-0.15		
Icel.	0.56	1.49	0.40	0.38	0.60	**0.51**	0.72	1.62	1.59	0.98	**2.46**	1.08	0.99	**2.64**	**1.31**	0.66	
Nital	**3.26**	**5.85**	**3.56**	**2.52**	**4.67**	**4.02**	**4.29**	**1.71**	**5.19**	2.45	**5.30**	3.98	**3.88**	3.10	**3.02**	2.23	**4.31**
Sard.	0.50	1.78	0.35	0.17	1.65	0.98	1.01	0.54	0.08	2.23	1.06	0.64	0.69	0.69	0.25	-0.19	1.16
Tusc.	-0.22	1.19	-0.17	-0.51	-0.12	-0.29	0.00	0.34	-0.42	0.83	1.16	-0.16	0.10	1.19	0.12	-0.53	0.02
Norw.	0.28	0.71	0.09	0.15	0.16	0.27	**0.91**	**2.19**	1.15	1.27	**0.90**	0.59	0.03	**2.45**	0.50	0.32	**0.77**
Port.	0.35	1.52	0.27	-0.16	0.41	0.36	1.18	1.59	0.73	2.60	1.61	-0.03	0.37	1.24	0.21	-0.21	1.07
Saa.	**17.63**	**19.01**	**17.54**	**18.69**	**20.18**	**18.22**	**18.88**	**18.00**	**16.91**	**20.01**	**12.94**	**17.76**	**17.85**	**14.50**	**17.11**	**16.61**	**19.04**
Syria	-0.37	1.49	-0.26	-0.59	-0.08	-0.02	0.62	-0.01	0.32	0.83	1.48	-0.15	0.07	0.27	0.02	-0.93	0.58
Spain	0.70	0.46	0.27	0.14	0.02	0.50	1.10	**2.64**	0.32	**2.80**	0.72	0.08	0.17	**2.48**	0.35	0.37	**1.25**
Swed.	**2.47**	2.36	**2.27**	2.70	**2.83**	**2.70**	**3.82**	**4.19**	**3.39**	**4.08**	1.54	2.17	**2.44**	2.19	**2.50**	1.91	**3.65**
Swi.	-0.12	1.76	0.69	1.06	0.75	0.96	**1.72**	2.62	1.37	**3.36**	1.40	0.00	0.93	1.11	1.51	0.66	1.63
Turk.	1.14	**3.63**	1.11	0.69	1.35	1.40	1.36	1.57	0.65	**2.78**	**2.09**	0.52	1.96	**2.28**	1.19	0.11	1.89
Irel.	-0.21	1.26	-0.13	-0.19	-0.07	0.02	0.28	1.40	1.20	1.57	**1.88**	0.31	0.41	1.35	0.71	0.37	0.10
Anc.	1.58	**4.08**	1.37	0.90	2.94	1.80	1.89	0.50	1.34	0.28	1.30	2.45	1.89	1.51	0.85	0.13	1.83
Bulg.	0.11	1.46	0.34	-0.31	0.92	0.72	0.75	0.54	0.29	2.10	0.74	0.64	0.36	0.75	0.25	-0.19	**1.35**
Arm.	0.38	**2.16**	0.57	-0.02	0.87	**1.00**	**1.07**	0.32	0.46	2.21	**1.58**	0.25	0.68	0.94	0.39	-0.37	**1.59**
Bel.	0.96	2.40	0.76	0.89	2.05	1.24	1.96	0.50	0.08	1.45	0.33	0.98	0.94	0.81	0.37	-0.43	1.89
Cze.	-0.23	0.92	-0.18	-0.29	0.12	-0.10	0.55	0.78	0.60	0.47	0.82	0.57	0.17	0.85	0.23	-0.56	0.25
Est.	0.23	1.07	0.36	0.38	0.70	0.53	**1.21**	1.46	0.50	1.29	0.56	0.73	0.24	1.25	0.40	0.04	**1.05**
Geo.	0.14	**2.20**	0.82	0.38	**1.68**	**1.37**	1.10	0.63	1.06	2.51	**2.00**	0.40	1.12	0.65	0.85	0.15	**1.56**
Gre.	0.59	**2.29**	0.58	0.25	0.95	**0.93**	0.85	1.35	0.22	2.83	1.09	0.68	0.70	1.67	0.50	0.38	**1.51**
Rome	1.73	**4.04**	1.99	1.74	**3.33**	**2.67**	1.65	0.24	0.29	3.35	1.68	1.42	**2.50**	0.64	1.45	0.71	**2.89**
Sic.	0.18	1.48	0.08	0.13	0.70	0.61	0.85	1.54	0.80	2.05	1.10	0.05	0.07	1.51	-0.10	0.51	0.85
Kar.	1.47	**2.42**	1.17	1.30	**2.55**	**1.83**	**2.48**	2.41	1.11	2.69	-0.34	1.53	0.65	1.81	0.37	0.77	**2.64**
Palest.	1.33	**3.24**	**1.41**	0.90	**1.74**	**1.79**	1.28	1.09	0.26	2.53	**2.00**	0.91	**1.79**	**1.99**	0.79	0.40	**2.10**
Pol.	-0.25	1.71	-0.35	0.02	-0.43	-0.24	0.64	1.94	0.83	1.79	1.29	-1.09	-0.30	1.33	-0.46	-0.45	0.54
Rom.	0.72	**3.00**	0.87	0.16	**2.05**	1.20	1.74	0.24	1.02	1.98	0.88	1.52	1.11	0.15	0.60	-0.63	**2.08**
Gypsy	**18.24**	**19.14**	**17.60**	**18.24**	**18.06**	**18.02**	**15.92**	**18.46**	**15.02**	**17.71**	**17.20**	**16.74**	**20.13**	**17.89**	**18.40**	**17.01**	**16.29**
Ady.	0.77	1.58	1.02	0.81	2.07	1.36	1.43	1.50	0.76	1.67	1.18	1.74	1.39	1.38	1.10	0.78	1.33
Oss.	1.58	**2.88**	**1.41**	1.36	2.15	**2.01**	**1.42**	1.56	0.01	2.18	0.85	1.42	**1.87**	**2.07**	1.05	0.87	**2.03**
WSib.	0.72	**3.72**	0.67	0.37	1.69	0.74	1.33	0.58	0.72	2.36	1.72	0.88	1.40	0.54	0.70	-0.48	1.31
EuRus.	0.51	**3.98**	0.86	0.74	1.54	1.11	1.30	-0.51	0.77	2.03	1.79	0.61	1.18	-0.20	0.91	-0.65	1.97
Ural.	**2.75**	**4.84**	2.98	**3.12**	**3.89**	**3.34**	**3.30**	3.24	0.61	**4.11**	0.83	2.62	3.12	**2.57**	2.55	2.14	**3.93**
Ukr.	-0.93	1.93	-0.72	-0.67	0.41	-0.50	-0.48	-1.24	-1.13	-0.63	-0.45	-0.23	-0.47	-0.67	-1.47	-1.25	-0.13
VFin.	0.38	3.62	0.64	0.40	1.75	0.49	1.49	0.42	1.11	-0.12	1.21	1.58	1.11	1.09	0.72	-0.64	1.05

(continued)

Table D.1 Matrix of F_{ST} values based on nucleotide sequences for the 51 populations. Significant values are indicated in bold type ($P = 0.01$, corrected for multiple comparisons). Values for the sample from the ancient population of Britain are shaded. NOTE– Table D.1 is divided in three parts, to follow in next page.

	Nital	Sard.	Tusc.	Norw.	Port.	Saa.	Syria	Spain	Swed.	Swi.	Turk.	Irel.	Anc.	Bulg.	Arm.	Bel.	Cze.
Sard.	2.93																
Tusc.	2.62	-0.08															
Norw.	4.62	1.26	0.22														
Port.	2.36	0.04	-0.48	0.84													
Saa.	**20.01**	**16.56**	**18.79**	**17.64**	**18.87**												
Syria	1.74	-0.38	-1.00	0.41	-0.69	18.20											
Spain	5.06	0.53	0.04	0.40	0.19	17.38	0.33										
Swed.	**5.71**	2.63	**2.92**	2.10	**2.77**	**9.28**	**2.21**	**2.12**									
Swi.	5.19	1.38	0.78	0.91	1.54	16.99	**0.17**	0.67	2.07								
Turk.	4.14	**0.96**	-0.05	2.26	1.02	17.85	0.43	1.06	3.87	**1.16**							
Irel.	3.71	0.68	-0.33	0.19	0.44	19.95	-0.51	0.30	2.52	0.61	**1.00**						
Anc.	1.77	**0.82**	0.62	1.77	1.33	17.28	0.81	2.57	3.76	3.57	1.78	1.98					
Bulg.	3.59	0.03	-0.07	0.73	0.20	16.38	-0.35	0.43	2.31	0.72	1.01	0.36	1.22				
Arm.	3.04	**-0.09**	-0.22	1.39	0.05	**15.78**	**-0.55**	0.75	2.89	0.74	**0.41**	0.53	1.68	0.12			
Bel.	3.21	-0.11	0.19	1.39	0.61	14.37	0.10	1.27	2.38	1.54	1.45	1.99	-0.22	0.38	0.52		
Cze.	2.78	0.36	-0.44	-0.12	0.27	17.56	-0.46	0.41	1.85	0.63	1.12	-0.38	0.17	0.13	0.68	0.36	
Est.	4.11	0.47	0.06	0.25	0.58	16.98	**0.16**	0.59	2.41	0.80	1.95	0.49	0.65	0.32	0.97	0.20	**-0.33**
Geo.	2.56	**-0.10**	0.16	1.62	**0.28**	**17.91**	-0.40	1.31	3.48	0.87	**0.96**	0.45	1.43	0.30	0.16	1.09	**0.60**
Gre.	4.61	**0.37**	-0.12	1.02	0.65	**17.56**	0.08	0.43	3.30	0.95	0.57	0.59	1.95	0.12	0.27	1.09	**0.69**
Rome	3.63	**-0.19**	0.45	3.20	**1.37**	**14.91**	0.79	1.99	3.54	2.01	0.83	2.14	**1.44**	0.94	0.47	0.31	**1.55**
Sic.	4.04	-0.04	-0.01	0.59	0.27	17.74	-0.26	0.11	2.47	0.67	1.11	0.29	2.19	0.12	-0.07	0.75	0.51
Kar.	4.36	**0.99**	1.40	1.02	**1.59**	**12.06**	**1.28**	1.12	1.22	1.85	2.39	2.28	0.91	0.72	1.17	-0.05	**1.07**
Palest.	3.72	**0.56**	**-0.33**	2.39	**0.64**	**17.68**	0.22	1.38	4.40	1.86	**-0.04**	1.23	**1.16**	**0.80**	0.45	0.88	**1.12**
Pol.	2.69	0.23	-0.52	0.03	-0.66	19.63	-1.11	-0.34	2.50	0.67	0.14	-0.53	1.50	0.33	-0.39	1.09	-0.24
Rom.	2.48	**0.19**	0.57	1.43	**0.57**	15.05	-0.13	1.35	2.20	1.35	1.24	1.19	0.36	-0.07	0.34	0.06	**0.20**
Gypsy	**21.14**	**17.21**	**15.20**	**19.47**	**18.20**	**26.88**	**16.97**	**16.71**	**16.17**	**18.13**	**12.48**	**17.45**	**17.11**	**17.66**	**16.30**	**16.57**	**17.13**
Ady.	3.69	0.35	0.14	1.31	0.93	17.89	0.29	1.32	2.88	2.09	2.00	1.22	0.64	0.37	1.23	0.21	0.23
Oss.	5.40	**0.78**	**0.40**	2.01	1.65	**15.14**	**1.21**	1.30	3.09	1.76	1.00	1.71	**1.13**	**0.84**	1.21	0.24	**1.07**
WSib.	2.70	**0.36**	-0.27	1.61	0.34	18.13	-0.65	1.15	3.10	1.45	-0.58	0.63	0.84	0.10	-0.25	0.49	0.43
EuRus.	2.16	**0.18**	0.32	1.81	0.11	16.16	-0.63	1.63	2.53	1.42	0.91	0.93	0.49	0.36	-0.16	0.13	0.48
Ural.	**6.77**	**2.45**	1.71	**3.16**	**2.98**	**13.47**	**2.40**	2.29	3.23	**2.33**	1.42	3.70	2.18	**1.81**	2.01	1.14	**2.39**
Ukr.	0.20	-1.22	-2.18	-0.44	-0.97	17.41	-1.96	-0.22	2.16	0.96	-0.90	-0.44	-2.19	-1.11	-1.16	-1.59	-1.78
VFin.	1.74	1.01	-0.25	0.74	0.99	17.54	-0.11	1.86	3.03	2.67	1.72	1.67	-1.00	0.83	0.65	-0.41	-0.16

(continued)

	Est.	Geo.	Gre.	Rome	Sic.	Kar.	Palest	Pol.	Rom.	Gypsy	Ady.	Oss.	WSib.	EuRus	Ural.	Ukr.
Geo.	1.02															
Gre.	**0.74**	0.82														
Rome	**1.72**	0.61	0.53													
Sic.	**0.53**	0.56	0.23	1.12												
Kar.	**0.61**	1.88	1.33	1.69	1.10											
Palest.	**1.41**	0.87	0.48	**0.32**	0.78	**2.11**										
Pol.	0.19	-0.05	0.35	1.27	0.03	**1.69**	0.29									
Rom.	**0.69**	0.53	0.84	**0.72**	1.16	**0.65**	1.20	0.82								
Gypsy	**18.56**	**18.06**	**17.47**	**14.98**	**17.37**	**18.76**	**13.59**	**17.83**	**18.87**							
Ady.	0.29	0.83	1.16	1.07	1.18	**1.69**	1.21	1.78	1.10	16.10						
Oss.	**1.16**	1.81	0.80	**0.41**	0.77	**1.27**	0.67	1.51	**1.32**	12.34	0.25					
WSib.	1.04	0.21	0.85	0.82	0.96	**1.79**	-0.23	0.53	-0.15	14.68	1.67	1.01				
EuRus.	0.98	0.24	1.25	0.37	0.91	**1.37**	0.80	-0.48	-0.21	17.96	1.77	1.51	-0.30			
Ural.	**1.97**	**2.98**	1.95	**1.51**	2.41	**1.16**	1.36	**3.29**	**2.01**	12.69	2.16	**0.34**	1.46	2.47		
Ukr.	-1.43	-1.17	-1.00	-1.43	-0.47	**-0.39**	-1.81	-1.04	-1.11	16.35	-0.99	-0.96	-1.00	-1.51	0.20	
VFin.	0.12	1.22	1.41	1.29	1.86	**1.19**	1.10	1.25	0.14	20.07	0.30	1.01	2.00	0.42	3.00	-1.63

	Aus	Basq	Eng	Corn	Wal	Scot	Wls	Skye	Ork	Den	Finl	Fran	Sge	Nge	CGer	Bav	Icel
Aus.	4.39	3.66	4.20	3.85	3.86	4.28	4.34	4.46	4.35	4.72	**4.24**	4.06	3.95	4.79	4.08	4.19	4.51
Basq.	0.04	2.85	3.47	3.13	3.11	**3.55**	**3.62**	**3.81**	3.61	**4.01**	**3.47**	3.34	3.20	**4.12**	3.34	3.49	3.79
Eng.	0.00	0.04	4.01	3.66	3.66	4.08	4.14	4.28	4.15	4.51	**4.03**	3.88	3.75	**4.62**	3.87	3.99	4.31
Corn.	0.00	0.04	-0.01	3.32	3.32	3.74	3.80	3.91	3.81	4.17	3.70	3.55	3.41	4.27	3.52	3.64	3.98
Wal.	0.01	0.03	0.01	0.01	3.31	3.73	3.81	3.98	3.82	4.19	3.72	3.53	3.41	**4.33**	3.55	3.67	3.98
Scot.	0.01	**0.05**	0.00	0.00	0.00	4.15	4.22	4.36	4.24	4.59	**4.13**	3.96	3.84	**4.71**	3.96	4.07	**4.39**
Wis.	0.02	**0.08**	0.02	0.02	0.04	0.03	4.23	4.38	4.27	4.65	**4.19**	4.03	**3.90**	**4.77**	4.03	4.16	4.44
Skye	0.04	**0.15**	0.05	0.02	0.09	0.06	0.04	4.46	4.41	4.76	4.33	4.20	4.05	4.79	4.15	4.25	4.60
Ork.	0.04	0.06	0.03	0.03	0.04	0.04	0.03	0.06	4.24	4.70	4.13	4.03	3.90	4.76	4.01	4.13	4.49
Den.	0.07	**0.12**	0.04	0.04	0.07	0.05	0.07	0.07	0.12	4.93	**4.56**	4.45	4.27	5.15	4.38	4.51	4.80
Finl.	**0.06**	**0.06**	**0.04**	0.05	0.08	**0.07**	**0.08**	0.11	0.02	**0.11**	3.98	3.91	3.76	**4.64**	3.87	4.01	**4.39**
Fran.	0.00	0.04	0.00	0.02	0.00	0.02	0.04	0.09	0.03	0.12	0.05	3.74	3.63	4.51	3.74	3.87	4.22
SGe.	0.01	0.03	0.00	0.00	0.01	0.02	**0.04**	0.07	0.04	0.06	0.03	0.01	3.50	**4.38**	3.60	3.75	**4.08**
NGe.	0.04	**0.14**	**0.07**	0.06	**0.12**	**0.08**	**0.10**	0.01	0.09	0.14	**0.10**	0.08	**0.08**	5.10	4.49	4.56	**4.97**
Cger.	0.03	0.06	0.01	0.00	0.04	0.03	0.05	0.07	0.04	0.06	0.02	0.01	0.00	0.08	3.71	3.85	**4.21**
Bav.	0.00	0.06	-0.01	-0.02	0.01	-0.01	0.05	0.02	0.01	0.05	0.03	0.00	0.00	0.01	-0.01	4.00	4.33
Icel.	0.03	0.07	0.02	0.02	0.03	**0.02**	0.03	0.08	0.07	0.04	**0.11**	0.05	**0.04**	**0.12**	**0.06**	0.04	4.58
Nital	**0.16**	**0.24**	**0.15**	0.11	**0.21**	**0.17**	**0.20**	0.10	**0.27**	0.14	**0.24**	0.20	**0.15**	0.17	**0.13**	0.12	**0.21**
Sard.	0.02	0.06	0.01	0.01	0.06	0.04	0.04	0.02	0.00	0.10	0.04	0.03	0.02	0.04	0.01	-0.01	0.05
Tusc.	-0.01	0.03	-0.02	-0.03	-0.01	-0.02	-0.01	0.02	-0.02	0.04	0.04	-0.01	-0.01	0.06	0.00	-0.02	0.00
Norw.	0.01	0.03	0.00	0.01	0.01	0.01	**0.03**	**0.08**	0.04	0.04	**0.03**	0.02	0.00	**0.10**	0.02	0.01	**0.03**
Port.	0.02	0.04	0.01	-0.01	0.01	0.02	0.05	0.06	0.03	0.11	0.07	0.00	0.01	0.06	0.01	-0.01	0.06
Saa.	**0.88**	**0.83**	**0.85**	**0.86**	**0.94**	**0.92**	**0.95**	**0.89**	**0.82**	**1.03**	**0.59**	**0.84**	**0.81**	**0.74**	**0.80**	**0.79**	**1.04**
Syria	-0.02	0.04	-0.02	-0.03	-0.01	-0.01	0.02	0.00	0.01	0.04	0.05	-0.01	-0.01	0.01	-0.01	-0.04	0.02
Spain	0.03	0.02	0.01	0.01	0.00	0.02	0.04	0.10	0.01	**0.11**	0.03	0.00	0.01	**0.11**	0.01	0.01	**0.06**
Swed.	**0.11**	0.08	**0.09**	0.10	**0.11**	**0.11**	**0.17**	**0.19**	**0.15**	**0.19**	0.06	0.09	**0.09**	0.11	0.10	0.08	**0.17**
Swi.	0.00	0.05	0.03	0.04	0.03	0.04	**0.07**	0.10	0.05	**0.13**	0.06	0.00	0.03	0.05	0.06	0.02	0.08
Turk.	0.05	**0.14**	0.04	0.03	0.06	0.05	0.06	0.09	0.03	**0.16**	**0.09**	0.03	0.07	**0.12**	0.05	0.01	0.08
Irel.	-0.01	0.04	0.00	-0.01	0.00	0.00	0.01	0.05	0.04	0.05	**0.07**	0.01	0.01	0.06	0.03	0.01	0.01
Anc.	0.07	**0.14**	0.05	0.03	0.11	0.06	0.08	0.02	0.06	0.01	0.04	0.11	0.06	0.08	0.02	0.01	0.08
Bulg.	0.00	0.05	0.01	-0.01	0.04	0.03	0.03	0.02	0.01	0.08	0.03	0.03	0.01	0.03	0.01	-0.01	**0.06**
Arm.	0.02	**0.10**	0.03	0.01	0.05	**0.04**	**0.05**	0.02	0.03	0.12	**0.08**	0.02	0.03	0.05	0.02	-0.01	**0.08**
Bel.	0.04	0.08	0.02	0.03	0.08	0.04	0.08	0.03	0.00	0.08	0.01	0.04	0.03	0.04	0.01	-0.02	0.08
Cze.	-0.01	0.03	-0.01	-0.01	0.00	0.00	0.02	0.03	0.03	0.02	0.03	0.02	0.00	0.04	0.01	-0.02	0.01
Est.	0.01	0.04	0.01	0.02	0.03	0.02	**0.05**	0.06	0.02	0.05	0.02	0.03	0.01	0.06	0.02	0.00	**0.05**
Geo.	0.01	**0.09**	0.03	0.02	**0.07**	**0.06**	0.05	0.03	0.05	0.12	**0.09**	0.02	0.04	0.03	0.04	0.01	**0.07**
Gre.	0.02	**0.08**	0.02	0.01	0.03	**0.04**	0.04	0.05	0.01	0.11	0.04	0.03	0.02	0.07	0.02	0.01	**0.07**
Rome	0.08	**0.14**	0.07	0.06	**0.13**	**0.10**	0.06	0.01	0.01	0.18	0.06	0.06	0.08	0.03	0.05	0.03	**0.13**
Sic.	0.01	0.05	0.01	0.00	0.02	0.03	0.04	0.05	0.03	0.08	0.04	0.00	0.00	0.07	0.00	0.02	0.04
Kar.	0.06	**0.08**	0.05	0.05	**0.09**	**0.08**	**0.10**	0.09	0.04	0.11	-0.01	0.06	0.02	0.08	0.01	0.03	**0.13**
Palest.	0.07	**0.15**	**0.06**	0.05	**0.09**	**0.07**	0.06	0.07	0.02	0.15	**0.09**	0.06	**0.08**	**0.11**	0.04	0.03	**0.10**
Pol.	0.00	0.05	0.00	0.00	-0.01	0.01	0.04	0.08	0.04	0.08	0.06	-0.03	0.00	0.08	-0.01	-0.01	0.05
Rom.	0.03	**0.11**	0.03	0.01	**0.08**	0.05	**0.08**	0.01	0.04	0.09	0.04	0.06	0.04	0.01	0.02	-0.02	**0.10**
Gypsy	**1.06**	**0.94**	**0.92**	**0.96**	**0.94**	**0.93**	**0.86**	**1.11**	**0.84**	**1.09**	**0.91**	**0.93**	**1.02**	**1.11**	**0.99**	**0.98**	**0.91**
Ady.	0.03	0.05	0.04	0.03	0.07	0.05	0.06	0.07	0.03	0.08	0.04	0.07	0.04	0.07	0.04	0.03	0.06
Oss.	0.08	**0.12**	0.06	0.06	0.10	**0.08**	**0.07**	0.09	0.00	0.12	0.04	0.07	**0.08**	**0.11**	0.05	0.05	**0.09**
WSib.	0.04	**0.11**	0.03	0.01	0.06	0.03	0.06	0.03	0.03	0.11	0.07	0.03	0.05	0.04	0.02	-0.02	0.07
EuRus.	0.02	0.13	0.02	0.02	0.05	0.03	0.05	-0.03	0.03	0.10	0.06	0.02	0.03	-0.01	0.02	-0.03	0.09
Ural.	**0.13**	**0.17**	0.12	0.12	**0.15**	**0.14**	**0.15**	0.16	0.03	**0.21**	0.03	0.11	0.11	**0.13**	0.10	0.10	**0.19**
Ukr.	-0.04	0.02	-0.04	-0.05	-0.01	-0.03	-0.03	-0.06	-0.05	-0.02	-0.03	-0.02	-0.04	-0.02	-0.07	-0.06	0.00
VFin.	0.02	0.10	0.02	0.00	0.05	0.02	0.06	0.02	0.05	0.00	0.04	0.06	0.03	0.06	0.02	-0.03	0.05

(continued)

Table D.2 Matrix of genetic distances (Nei, 1987) based on nucleotide sequences and assuming Tamura-Nei model of nucleotide evolution, with $\alpha = 0.26$ (Meyer *et al.*, 1999) for the 51 populations. Values above the diagonal are genetic distances between populations (d_{XY}) and below the diagonal are the net genetic distances (d_A). Values in the diagonal (showed with doted line) are internal nucleotide diversity (d_X). Significant values are indicated in bold type ($P = 0.01$, corrected for multiple comparisons). Values for the sample from the ancient population of Britain are shaded. NOTE– Table D.2 is divided in three parts, to follow in next two pages.

	Nital	Sard	Tus	Norw	Port	Saa	Syria	Sp	Swe	Swi	Turk	Irel	Anc	Bulg	Arm	Bel	Cze
Aus.	**5.29**	4.40	4.69	4.09	3.98	**5.05**	4.77	4.13	**4.46**	3.90	5.05	3.86	5.06	4.20	4.90	4.89	4.30
Basq.	**4.61**	3.67	3.97	3.35	3.24	**4.23**	4.07	3.35	3.66	3.19	4.38	3.14	4.37	3.48	4.22	4.16	3.57
Eng.	**5.09**	4.20	4.50	3.90	3.79	**4.82**	4.58	3.92	**4.25**	3.74	4.85	3.67	4.85	4.02	4.72	4.68	4.11
Corn.	4.71	3.85	4.14	3.56	3.43	**4.50**	4.23	3.57	3.92	3.41	4.50	3.33	4.49	3.65	4.36	4.35	3.77
Wal.	**4.80**	3.90	4.15	3.55	3.44	**4.57**	4.24	3.56	**3.92**	3.39	4.52	3.32	4.57	3.69	4.39	4.39	3.77
Scot.	**5.18**	4.30	4.56	3.98	3.87	**4.97**	4.66	4.00	**4.35**	3.83	4.93	3.75	4.94	4.11	4.81	4.77	4.19
Wis.	**5.26**	4.34	4.62	**4.04**	3.94	**5.04**	4.74	4.07	**4.44**	3.90	4.98	3.80	4.99	4.15	4.86	4.86	4.26
Skye	5.26	4.44	4.76	**4.20**	4.07	**5.09**	4.83	4.24	**4.58**	4.04	5.12	3.95	5.06	4.25	4.94	4.91	4.38
Ork.	**5.33**	4.30	4.61	4.05	3.92	**4.91**	4.73	4.04	**4.43**	3.89	4.96	3.84	4.98	4.13	4.84	4.78	4.26
Den.	5.54	4.75	5.01	4.39	4.34	**5.47**	5.11	**4.48**	**4.82**	4.31	5.43	4.19	5.28	4.55	5.28	5.20	4.60
Finl.	**5.17**	4.21	4.54	**3.91**	3.82	**4.55**	4.64	3.92	4.21	3.76	4.88	3.73	4.83	4.02	4.75	4.65	4.14
Fran.	5.01	4.08	4.37	3.78	3.64	**4.69**	4.46	3.78	4.12	3.58	4.71	3.55	4.78	3.90	4.58	4.57	4.01
SGe.	**4.84**	3.95	4.25	3.64	3.53	**4.53**	4.34	3.66	**3.99**	3.49	4.63	3.44	4.61	3.76	4.47	4.43	3.87
NGe.	5.66	4.77	5.12	**4.54**	4.39	**5.27**	5.17	**4.56**	4.82	4.32	5.48	4.28	5.43	4.59	5.29	5.25	4.71
Cger.	**4.93**	4.05	4.36	3.77	3.64	**4.63**	4.45	3.77	4.11	3.63	4.71	3.55	4.68	3.87	4.57	4.52	3.98
Bav.	5.06	4.17	4.48	3.90	3.76	**4.76**	4.56	3.92	4.24	3.73	4.82	3.68	4.80	3.99	4.68	4.64	4.09
Icel.	**5.43**	4.53	4.80	**4.21**	4.12	**5.30**	4.91	**4.25**	4.62	4.08	5.18	3.97	5.17	4.36	5.06	5.03	4.42
Nital	5.87	5.27	5.60	**5.01**	4.83	**6.02**	5.64	5.06	5.41	4.90	5.99	4.77	5.84	5.10	5.80	5.78	5.19
Sard.	0.15	4.36	4.68	4.12	3.96	**4.96**	4.76	4.10	4.46	3.95	5.04	3.88	5.02	4.18	4.87	4.83	4.31
Tusc.	0.15	-0.01	5.02	4.40	4.26	**5.44**	5.06	4.41	4.81	4.25	5.32	4.16	5.34	4.50	5.19	5.18	4.60
Norw.	**0.18**	0.05	0.00	3.78	3.69	**4.68**	4.49	3.81	4.13	3.64	4.78	3.57	4.74	3.92	4.64	4.59	4.00
Port.	0.12	0.00	-0.02	0.03	3.54	**4.64**	4.34	3.68	4.04	3.54	4.63	3.46	4.63	3.78	4.48	4.46	3.90
Saa.	**1.11**	**0.80**	**0.96**	**0.82**	**0.90**	3.95	**5.50**	**4.69**	**4.54**	**4.46**	**5.74**	**4.58**	**5.66**	**4.75**	**5.54**	**5.34**	**4.95**
Syria	0.10	-0.02	-0.05	0.00	-0.03	**0.93**	5.20	4.51	4.86	4.31	5.43	4.24	5.45	4.58	5.26	5.26	4.69
Spain	0.22	0.02	-0.01	0.02	0.01	**0.82**	0.00	3.81	4.15	3.64	4.75	3.59	4.80	3.92	4.63	4.60	4.04
Swed.	**0.31**	0.12	0.14	**0.08**	0.11	**0.41**	0.10	**0.08**	4.32	3.95	5.17	3.92	5.15	4.26	5.01	4.93	4.35
Swi.	**0.25**	0.05	0.03	0.04	0.05	**0.78**	0.00	0.03	0.08	3.43	4.57	3.40	4.66	3.74	4.45	4.43	3.85
Turk.	**0.25**	0.05	0.00	**0.08**	0.05	**0.96**	0.03	0.04	**0.20**	0.05	5.61	4.52	5.71	4.85	5.52	5.55	4.98
Irel.	**0.16**	0.02	-0.02	0.01	0.01	**0.93**	-0.03	0.01	**0.09**	0.02	0.04	3.34	4.54	3.69	4.39	4.40	3.77
Anc.	0.10	0.04	0.03	0.05	0.06	**0.89**	0.04	0.10	0.19	0.15	0.10	0.07	5.60	4.85	5.58	5.44	4.92
Bulg.	**0.16**	0.00	-0.01	0.03	0.01	**0.78**	-0.02	0.02	**0.10**	0.03	0.04	0.01	0.04	4.00	4.70	4.67	4.12
Arm.	**0.17**	0.00	-0.01	**0.06**	0.01	**0.88**	-0.03	**0.04**	**0.16**	0.04	0.02	0.03	0.09	0.01	5.38	5.37	4.84
Bel.	**0.19**	-0.01	0.01	0.04	0.03	**0.71**	0.01	0.04	0.12	0.06	0.08	0.07	-0.01	0.01	0.03	5.31	4.79
Cze.	0.14	0.02	-0.02	-0.01	0.01	**0.86**	-0.02	0.01	0.08	0.02	0.05	-0.01	0.00	0.00	0.04	0.01	4.23
Est.	**0.19**	0.02	0.00	0.01	0.03	**0.83**	0.00	0.02	**0.10**	0.03	0.09	0.02	0.02	0.01	0.05	0.00	-0.01
Geo.	0.13	0.00	0.01	**0.06**	0.02	**0.93**	-0.02	**0.05**	**0.17**	0.04	0.05	0.02	0.07	0.01	0.01	0.05	0.03
Gre.	**0.21**	0.01	-0.01	**0.04**	0.03	**0.82**	-0.01	0.02	**0.13**	0.04	0.02	0.02	0.07	0.00	0.02	0.04	0.03
Rome	**0.22**	-0.01	0.02	0.11	0.06	**0.74**	0.04	0.07	**0.18**	0.08	0.05	0.08	0.08	0.03	0.02	0.02	0.07
Sic.	**0.18**	0.00	-0.01	0.02	0.01	**0.81**	-0.02	0.01	**0.09**	0.02	0.05	0.01	0.08	0.01	0.00	0.02	0.02
Kar.	**0.21**	0.04	0.05	0.04	0.06	**0.53**	0.05	0.04	0.05	0.07	0.11	**0.08**	0.03	0.03	**0.06**	-0.01	0.04
Palest.	**0.23**	0.03	-0.01	**0.09**	0.05	**1.01**	0.02	**0.06**	**0.25**	0.10	0.00	0.06	0.07	0.04	0.02	0.05	0.06
Pol.	0.15	0.02	-0.01	0.01	-0.02	**0.93**	-0.04	0.00	0.10	0.03	0.03	-0.01	0.08	0.02	0.01	0.06	0.00
Rom.	**0.12**	0.01	0.02	**0.05**	0.03	**0.71**	-0.01	0.05	0.10	0.05	0.06	0.04	0.01	0.00	0.02	0.00	0.01
Gypsy	**1.45**	**1.00**	**0.91**	**0.96**	**1.03**	**1.61**	**1.05**	**0.86**	**0.93**	**0.98**	**0.76**	**0.90**	**1.08**	**0.96**	**1.03**	**1.03**	**0.98**
Ady.	0.21	0.02	0.01	0.04	0.04	**0.89**	0.01	0.05	0.13	0.08	0.11	0.04	0.03	0.01	0.07	0.01	0.01
Oss.	**0.32**	0.04	0.02	**0.08**	0.09	**0.80**	0.07	0.05	**0.16**	**0.09**	0.06	0.08	0.06	0.04	**0.07**	0.01	0.05
WSib.	0.16	0.02	-0.01	0.06	0.01	**0.87**	-0.03	0.04	0.13	0.05	-0.02	0.02	0.05	0.00	0.00	0.03	0.02
EuRus.	0.13	0.00	0.02	0.05	0.00	**0.79**	-0.03	0.05	0.12	0.05	0.05	0.02	0.03	0.00	-0.01	0.01	0.02
Ural.	**0.40**	**0.11**	0.09	**0.12**	0.13	**0.64**	0.12	**0.09**	**0.15**	0.09	0.08	**0.14**	0.12	0.07	**0.11**	0.06	0.11
Ukr.	0.04	-0.06	-0.10	-0.04	-0.05	**0.83**	-0.08	-0.03	0.09	0.01	-0.02	-0.04	-0.09	-0.06	-0.04	-0.06	-0.08
VFin.	0.11	0.04	-0.01	0.02	0.03	**0.85**	0.00	0.07	0.14	0.09	0.10	0.05	-0.04	0.03	0.05	-0.01	-0.01

(continued)

	Est	Geo	Gre	Rome	Sic	Kar	Palest	Pol	Rom	Gyp	Ady	Oss	wSib	ERus	Ural	Ukr	Vfin
Aus.	4.28	4.58	4.09	5.14	3.91	4.07	5.22	3.60	4.32	5.80	4.53	5.00	4.17	4.82	4.77	4.40	4.44
Basq.	3.55	3.89	3.38	4.44	3.18	3.32	4.54	2.88	3.63	4.93	3.78	4.28	3.48	4.16	4.05	3.70	3.76
Eng.	4.10	4.41	3.90	4.95	3.72	3.87	5.02	3.41	4.13	5.48	4.35	4.79	3.97	4.64	4.57	4.21	4.25
Corn.	3.76	4.06	3.54	4.60	3.37	3.52	4.67	3.07	3.77	5.18	3.99	4.45	3.61	4.29	4.23	3.86	3.89
Wal.	3.77	4.10	3.56	4.66	3.39	3.56	4.70	3.05	3.83	5.15	4.04	4.48	3.65	4.31	4.26	3.89	3.94
Scot.	4.18	4.51	3.99	5.05	3.81	3.97	5.11	3.49	4.22	5.56	4.43	4.88	4.05	4.72	4.66	4.29	4.32
Wis.	4.25	4.54	4.02	5.05	3.86	4.04	5.14	3.56	4.29	5.53	4.48	4.91	4.12	4.78	4.71	4.34	4.41
Skye	4.37	4.64	4.15	5.11	3.99	4.14	5.26	3.72	4.34	5.90	4.61	5.05	4.20	4.81	4.84	4.42	4.48
Ork.	4.22	4.55	4.00	5.00	3.86	3.98	5.10	3.57	4.26	5.52	4.46	4.85	4.09	4.76	4.60	4.31	4.40
Den.	4.60	4.96	4.45	5.52	4.25	4.39	5.58	3.95	4.65	6.11	4.85	5.31	4.52	5.18	5.12	4.69	4.69
Finl.	4.09	4.45	3.90	4.92	3.74	3.79	5.04	3.45	4.12	5.46	4.34	4.75	4.00	4.66	4.47	4.21	4.26
Fran.	3.98	4.27	3.77	4.80	3.58	3.74	4.89	3.24	4.03	5.36	4.25	4.67	3.85	4.50	4.43	4.10	4.16
SGe.	3.84	4.17	3.65	4.70	3.46	3.59	4.79	3.15	3.89	5.32	4.10	4.55	3.74	4.39	4.31	3.96	4.01
NGe.	4.69	4.96	4.50	5.45	4.33	4.45	5.62	4.04	4.66	6.22	4.93	5.39	4.53	5.15	5.13	4.78	4.84
Cger.	3.95	4.27	3.75	4.78	3.56	3.69	4.86	3.25	3.98	5.40	4.20	4.63	3.82	4.49	4.41	4.03	4.11
Bav.	4.08	4.39	3.88	4.90	3.72	3.84	4.99	3.39	4.07	5.54	4.34	4.78	3.92	4.58	4.55	4.19	4.20
Icel.	4.42	4.74	4.23	5.29	4.04	4.23	5.35	3.74	4.49	5.76	4.66	5.11	4.30	4.99	4.93	4.54	4.57
Nital	5.21	5.44	5.02	6.03	4.83	4.96	6.13	4.49	5.16	6.94	5.45	5.98	5.03	5.67	5.78	5.22	5.27
Sard.	4.28	4.55	4.06	5.04	3.89	4.04	5.18	3.60	4.29	5.74	4.50	4.95	4.14	4.79	4.75	4.37	4.45
Tusc.	4.59	4.89	4.37	5.40	4.21	4.38	5.46	3.90	4.63	5.98	4.82	5.26	4.44	5.13	5.05	4.66	4.73
Norw.	3.98	4.33	3.80	4.87	3.62	3.74	4.94	3.31	4.04	5.40	4.24	4.69	3.89	4.55	4.46	4.10	4.14
Port.	3.88	4.17	3.67	4.70	3.49	3.65	4.78	3.16	3.89	5.35	4.12	4.59	3.72	4.38	4.35	3.97	4.03
Saa.	**4.88**	**5.28**	**4.67**	**5.59**	**4.50**	**4.32**	**5.95**	**4.31**	**4.78**	**6.14**	**5.17**	**5.50**	**4.79**	**5.38**	**5.07**	**5.05**	**5.05**
Syria	4.68	4.95	4.47	5.52	4.29	4.47	5.58	3.97	4.69	6.21	4.92	5.40	4.52	5.18	5.17	4.77	4.83
Spain	4.01	4.33	3.79	4.85	3.62	3.76	4.93	3.31	4.05	5.32	4.26	4.69	3.89	4.57	4.44	4.13	4.20
Swed.	4.34	4.70	4.16	5.21	3.96	4.02	5.37	3.67	4.35	5.65	4.60	5.05	4.23	4.89	4.76	4.50	4.52
Swi.	3.83	4.13	3.62	4.66	3.44	3.60	4.77	3.14	3.86	5.25	4.10	4.53	3.70	4.37	4.25	3.97	4.03
Turk.	4.98	5.23	4.70	5.73	4.56	4.73	5.77	4.24	4.96	6.12	5.22	5.59	4.73	5.47	5.34	5.03	5.14
Irel.	3.77	4.07	3.56	4.62	3.39	3.57	4.69	3.06	3.81	5.13	4.02	4.48	3.63	4.30	4.26	3.88	3.95
Anc.	4.90	5.24	4.74	5.76	4.59	4.65	5.83	4.28	4.91	6.44	5.14	5.59	4.79	5.44	5.37	4.95	4.98
Bulg.	4.10	4.39	3.88	4.90	3.72	3.85	5.00	3.43	4.10	5.52	4.32	4.77	3.95	4.62	4.52	4.19	4.26
Arm.	4.82	5.07	4.58	5.58	4.40	4.57	5.67	4.10	4.81	6.27	5.07	5.48	4.63	5.29	5.25	4.90	4.97
Bel.	4.74	5.08	4.57	5.55	4.39	4.47	5.67	4.12	4.75	6.24	4.97	5.40	4.63	5.27	5.17	4.84	4.87
Cze.	4.18	4.52	4.01	5.06	3.84	3.97	5.14	3.52	4.22	5.65	4.43	4.90	4.08	4.74	4.67	4.28	4.34
Est.	4.16	4.50	3.98	5.02	3.81	3.92	5.11	3.51	4.21	5.67	4.40	4.86	4.07	4.72	4.61	4.26	4.31
Geo.	0.05	4.75	4.28	5.27	4.11	4.28	5.38	3.80	4.50	6.01	4.72	5.20	4.33	4.99	4.97	4.58	4.67
Gre.	0.03	0.04	3.74	4.75	3.59	3.74	4.85	3.30	4.00	5.35	4.22	4.63	3.84	4.52	4.40	4.06	4.15
Rome	0.07	0.02	0.01	5.74	4.62	4.75	5.85	4.34	5.00	6.36	5.23	5.62	4.86	5.50	5.40	5.07	5.18
Sic.	0.02	0.03	0.01	0.04	3.42	3.56	4.71	3.12	3.85	5.18	4.06	4.47	3.68	4.34	4.25	3.92	4.00
Kar.	0.03	0.08	0.05	0.07	0.04	3.63	4.89	3.28	3.94	5.40	4.19	4.60	3.82	4.47	4.31	4.03	4.08
Palest.	0.07	0.05	0.02	0.02	0.04	**0.11**	5.92	4.41	5.12	6.39	5.34	5.73	4.91	5.62	5.49	5.14	5.27
Pol.	0.02	0.02	0.02	0.07	0.01	0.06	0.04	2.81	3.54	4.96	3.79	4.23	3.36	4.00	4.00	3.59	3.68
Rom.	0.03	0.03	0.03	0.03	0.04	0.03	0.06	0.04	4.19	5.74	4.45	4.89	4.03	4.69	4.64	4.29	4.33
Gypsy	**1.04**	**1.08**	**0.92**	**0.93**	**0.91**	**1.03**	**0.87**	**1.00**	**1.09**	5.11	5.82	6.03	5.33	6.29	5.74	5.80	6.04
Ady.	0.01	0.04	0.04	0.06	0.04	0.06	0.08	0.08	0.05	**0.95**	4.61	5.05	4.32	5.00	4.86	4.51	4.55
Oss.	0.05	**0.09**	0.04	0.02	0.04	0.06	0.04	0.10	0.07	**0.74**	0.02	5.45	4.73	5.42	5.20	4.95	5.02
WSib.	0.05	0.02	0.03	0.05	0.03	0.07	0.01	0.02	0.00	**0.83**	0.08	0.06	3.88	4.54	4.46	4.14	4.25
EuRus.	0.03	0.01	0.04	0.02	0.02	0.05	0.05	-0.02	-0.02	**1.12**	0.09	0.08	-0.01	5.22	5.19	4.79	4.86
Ural.	**0.08**	**0.15**	0.07	0.08	0.09	0.04	0.08	0.15	0.09	**0.73**	0.11	0.02	0.07	0.13	4.90	4.72	4.83
Ukr.	-0.07	-0.05	-0.06	-0.05	-0.04	-0.03	-0.07	-0.06	-0.05	**1.00**	-0.04	-0.03	-0.05	-0.06	0.02	4.50	4.41
VFin.	0.00	0.06	0.05	0.08	0.06	0.04	0.08	0.04	0.00	**1.25**	0.01	0.06	0.08	0.02	0.15	-0.07	4.46

hg	Ancient	Austria	Adygei	Armenia	Basques	Belarus	Bulgaria	England	Cornwall	Wales	Denmark	Ireland	Scotland	Is.Skye	Western I	Orkney	Czech	Estonia	Finland	France	Georgia	south G	north G	Bava	cent G	Greece
A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B	-	-	-	0.005	-	-	-	-	-	-	-	-	0.001	-	-	-	-	-	-	-	0.008	-	-	-	-	-
C	-	-	0.060	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.016	-	-	-	-	-	-
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H*	0.313	0.552	0.560	0.471	0.686	0.509	0.518	0.523	0.570	0.609	0.458	0.554	0.492	0.388	0.365	0.410	0.482	0.584	0.449	0.594	0.492	0.620	0.467	0.490	0.564	0.488
I	0.021	0.034	-	0.016	-	0.036	0.028	0.023	0.047	0.033	0.042	0.030	0.042	0.020	0.072	0.064	0.036	0.007	0.028	-	0.024	0.030	0.009	0.020	0.026	0.040
IXW	0.042	-	0.020	0.005	-	0.018	0.028	0.008	-	-	-	-	0.006	-	0.011	-	-	0.013	0.023	0.047	0.032	0.005	0.028	-	0.034	0.016
J1	-	0.009	-	0.016	-	-	0.014	0.004	-	0.011	-	-	0.006	-	0.017	-	0.012	-	0.006	-	-	0.015	-	-	-	-
J1a	-	0.026	-	0.005	0.010	0.018	-	0.023	-	-	0.063	-	0.004	-	-	-	0.012	0.007	-	-	-	0.010	0.037	0.041	0.009	-
J1b	-	-	-	0.031	-	-	-	-	-	-	-	-	0.001	-	-	-	0.012	-	-	-	0.008	0.005	-	-	-	-
J1b1	-	-	-	-	-	0.018	-	0.012	0.012	0.033	-	-	0.035	-	0.011	0.038	-	0.013	0.006	0.016	-	-	-	-	-	-
J2	-	-	0.020	0.021	-	-	-	0.004	-	-	-	0.010	0.012	-	0.022	-	-	-	0.011	0.031	0.016	-	-	-	0.009	0.008
J*	0.083	0.069	0.020	0.031	0.019	0.073	0.071	0.078	0.151	0.109	0.063	0.050	0.088	0.102	0.105	0.077	0.048	0.067	0.051	0.031	0.024	0.060	0.056	0.102	0.034	0.152
K	0.021	0.103	0.020	0.073	0.048	0.018	0.064	0.066	0.035	0.076	0.021	0.069	0.067	0.143	0.133	0.090	0.036	0.034	0.028	0.063	0.105	0.050	0.093	0.020	0.034	0.040
L	-	-	-	-	-	0.018	0.007	0.012	-	-	-	-	0.002	-	-	0.026	-	-	-	0.016	-	-	0.009	-	-	-
N	-	-	-	0.058	-	0.018	0.007	-	-	-	-	-	-	-	0.006	-	-	0.013	-	-	0.024	0.005	-	-	-	0.016
T	-	0.034	0.080	0.037	0.019	0.018	0.007	0.019	0.058	0.011	-	0.010	0.016	-	0.061	0.026	0.036	0.027	-	0.031	0.065	0.015	0.009	0.041	0.026	0.008
T1	0.083	0.017	-	0.052	-	0.036	0.064	0.023	0.023	0.022	0.063	0.020	0.027	0.163	-	0.038	0.036	0.007	0.017	-	0.040	0.015	0.056	0.041	0.026	0.016
T2	0.104	0.026	0.020	0.021	0.010	-	0.014	0.039	-	0.011	0.104	0.059	0.051	0.041	0.033	-	0.036	0.034	0.040	0.047	0.008	0.045	0.009	0.041	0.051	0.008
T3	-	-	0.020	0.005	0.010	-	0.014	0.004	-	-	-	-	0.001	-	0.011	-	0.024	0.007	-	-	0.032	-	0.009	-	0.009	0.008
T4	0.021	0.009	0.020	-	0.010	0.018	0.007	0.004	-	-	0.021	-	0.001	-	-	-	-	-	0.006	-	-	0.005	0.019	-	0.009	0.008
T5	0.021	0.009	-	-	0.010	0.018	-	0.008	0.012	-	-	-	0.002	-	0.006	-	-	-	-	-	-	-	-	-	0.009	0.008
U1	-	0.009	0.060	0.052	-	-	0.007	0.008	-	-	-	0.010	0.001	-	0.028	0.013	0.012	-	0.006	0.016	0.040	0.005	-	-	0.017	0.024
U2	-	0.009	0.020	0.010	-	-	0.007	0.012	-	-	-	0.010	0.008	-	0.006	-	-	0.007	0.011	0.016	-	0.005	0.009	-	0.009	0.008
U5	-	-	-	0.005	0.010	-	-	0.008	-	0.022	0.021	0.010	0.010	-	-	0.038	0.012	-	-	-	-	0.010	-	-	-	-
U5a	-	-	-	-	0.095	0.018	-	0.016	-	-	-	-	0.010	-	-	-	0.012	0.007	0.045	0.016	-	-	0.009	-	0.009	-
U5a1	0.083	0.060	0.060	0.005	-	0.055	0.050	0.027	0.012	0.011	0.063	0.020	0.021	0.020	0.022	0.051	0.060	0.074	0.034	-	0.024	0.025	0.037	-	0.026	0.032
U5a1a	0.042	0.009	0.020	0.016	-	-	-	0.004	0.023	0.011	0.042	0.020	0.021	0.082	0.022	-	0.024	0.013	0.040	-	-	0.030	0.028	-	0.009	0.024
U5b	0.042	-	-	-	-	0.036	0.028	0.012	0.023	-	-	0.010	0.008	-	0.006	0.026	0.012	0.020	0.023	-	0.016	0.025	0.019	0.082	0.017	0.016
U5b1	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	0.034	-	-	-	-	-	-	-
U6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
U7	-	-	-	0.010	-	0.018	0.021	-	-	-	-	-	0.002	-	-	-	-	-	-	-	0.008	-	-	-	0.009	0.016
V	0.042	0.009	-	-	0.067	0.018	0.021	0.039	0.023	0.033	0.021	0.079	0.040	-	0.028	-	0.060	0.020	0.080	0.047	0.008	0.010	0.065	0.061	0.034	-
W	0.042	0.009	-	0.010	-	0.018	0.007	0.012	0.012	-	0.021	0.030	0.019	-	0.022	0.013	-	0.034	0.051	-	0.016	0.010	0.009	0.041	0.026	0.016
X	0.042	0.009	-	0.026	0.010	0.018	0.014	0.008	-	0.011	-	0.010	0.004	0.041	0.017	0.090	0.036	0.007	0.011	0.016	-	-	-	0.020	0.009	0.048
Z	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
other	-	-	-	0.005	-	-	-	0.008	-	-	-	-	0.001	-	-	-	-	-	-	-	0.008	-	0.019	-	-	-

(continued)

Table E.1 Haplogroup frequency distribution for the 51 populations. Haplogroup are defined according to the algorithm described in Chaper II.3.3.3

hg	Iceland	north It	Tuscany	Sardinia	Sicily	Rome	Karelia	Norway	Palestine	Poland	Portugal	Romania	Gypsy	Uralic	Ossetia	VolgaFinr	Siberia	E.Rus	Ukraine	Saami	Syria	Spain	Sweden	Switz.	Turkey
A	-	-	-	-	-	-	-	0.002	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B	-	-	-	-	-	-	-	-	-	-	-	0.011	-	-	-	-	-	-	-	-	-	-	-	-	-
C	-	-	-	-	-	-	-	-	-	-	-	-	-	0.019	0.028	-	-	-	-	-	-	-	-	-	0.028
D	-	-	-	-	-	-	0.012	-	-	-	-	-	-	-	-	-	-	-	-	0.051	-	-	-	-	-
H*	0.482	0.412	0.469	0.551	0.600	0.250	0.518	0.531	0.385	0.622	0.593	0.446	0.243	0.340	0.340	0.438	0.474	0.475	0.444	0.057	0.490	0.580	0.400	0.632	0.389
I	0.047	-	0.041	0.014	0.011	0.042	0.012	0.018	0.017	-	-	-	0.542	0.019	0.047	0.031	0.026	0.025	-	-	-	0.017	-	0.013	0.028
IXW	0.004	-	-	-	-	-	0.036	0.007	0.060	0.027	0.037	-	0.009	0.189	0.038	-	0.158	0.025	0.111	0.006	-	0.011	-	0.013	0.097
J1	0.002	-	-	-	0.011	0.021	-	0.009	0.009	0.027	-	0.043	0.065	-	0.009	-	0.026	-	-	-	-	0.006	-	0.013	0.014
J1a	0.009	-	0.020	-	-	-	-	0.009	0.017	-	-	-	-	0.019	0.009	-	-	-	-	-	0.020	0.006	0.017	0.039	0.014
J1b	-	-	0.020	-	-	-	-	-	0.017	-	-	-	-	-	-	-	-	0.025	-	-	-	0.011	-	-	0.014
J1b1	0.049	-	-	-	-	-	-	0.016	-	-	-	-	-	0.019	-	0.031	-	-	-	-	-	0.006	-	-	0.028
J2	0.013	-	0.041	0.029	0.022	0.021	-	-	0.026	-	-	-	-	-	0.057	0.063	-	-	-	-	-	0.017	-	-	0.042
J*	0.066	0.059	0.102	0.029	0.022	0.042	0.048	0.071	0.060	0.081	0.074	0.065	-	0.038	0.028	0.094	0.026	0.075	0.056	-	0.041	0.055	0.033	0.039	0.069
K	0.081	0.074	0.061	0.043	0.067	0.104	0.024	0.058	0.068	0.081	0.074	0.033	0.028	0.019	0.028	0.031	0.026	0.100	0.056	-	0.082	0.039	0.017	0.066	0.028
L	-	-	0.020	0.014	0.011	0.042	-	0.004	0.051	-	-	-	-	-	0.009	-	-	-	-	-	-	0.017	-	-	0.014
N	-	-	-	0.014	0.022	-	-	0.007	0.034	-	-	-	-	0.038	-	-	-	-	-	-	0.020	-	-	-	0.014
T	0.062	0.162	0.020	0.014	0.056	0.021	-	0.016	0.034	-	0.019	0.011	0.019	0.019	0.047	-	0.079	-	-	-	0.020	0.006	-	-	0.056
T1	0.004	0.103	0.020	0.029	0.022	0.021	0.036	0.007	0.034	0.027	0.037	0.087	0.028	0.019	0.019	0.031	0.079	0.100	0.056	-	0.061	0.017	0.033	-	0.028
T2	0.009	-	0.020	-	-	0.063	0.024	0.042	0.017	0.027	0.037	0.022	-	0.019	0.009	0.094	0.026	0.050	0.111	-	0.041	0.022	0.033	0.026	-
T3	-	0.029	-	0.058	-	-	-	0.004	0.017	0.027	0.019	-	-	-	0.009	0.031	-	0.025	-	-	-	0.006	-	-	-
T4	-	-	0.020	-	-	-	0.012	0.004	0.009	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.014
T5	0.028	-	0.020	0.014	-	-	-	0.011	0.009	-	-	0.011	-	-	-	-	-	-	-	-	0.020	-	-	-	-
U1	0.039	-	-	0.058	0.022	0.083	-	0.007	0.009	-	-	-	-	-	0.066	-	-	-	-	-	0.020	0.006	-	-	0.014
U2	-	0.015	-	-	0.011	0.021	-	0.002	0.009	-	-	-	-	0.038	0.019	-	-	-	-	-	0.020	0.006	-	0.013	-
U5	0.004	0.029	0.020	0.014	0.011	0.021	-	0.019	-	-	-	-	0.009	-	-	-	0.053	-	-	-	-	-	-	0.013	-
U5a	-	-	-	0.058	-	-	0.036	0.014	-	-	-	0.022	-	-	-	-	-	-	-	-	-	0.017	-	-	-
U5a1	0.047	0.015	0.041	-	-	0.021	0.024	0.042	-	-	-	0.033	-	0.057	0.075	0.125	-	-	0.056	0.011	0.020	0.022	0.033	0.013	-
U5a1a	0.009	0.015	-	-	0.011	0.021	0.048	0.011	0.009	-	-	0.022	0.009	0.019	0.009	-	-	0.025	0.056	-	0.020	0.006	-	0.039	-
U5b	0.013	0.029	-	-	-	0.063	0.012	0.018	-	-	-	0.054	-	0.019	-	-	-	0.025	-	0.017	-	0.006	-	0.013	-
U5b1	-	-	-	-	-	-	0.060	0.009	-	-	-	-	-	0.019	-	-	-	-	-	0.426	-	-	0.133	-	-
U6	-	-	-	-	0.011	-	-	-	0.009	-	0.056	-	-	-	-	-	-	-	-	-	-	0.011	-	-	-
U7	-	-	-	0.014	0.022	-	-	-	0.017	-	-	-	-	-	-	-	-	-	-	-	0.041	-	0.017	-	-
V	0.015	0.044	-	0.014	0.033	0.042	0.060	0.035	-	0.027	0.037	0.065	0.019	0.038	0.038	0.031	0.026	0.025	-	0.398	0.020	0.061	0.267	0.053	-
W	-	-	0.020	0.014	-	0.042	0.036	0.018	0.017	0.027	-	0.065	0.009	-	0.028	-	-	0.025	-	-	0.020	0.039	-	0.013	0.056
X	0.015	-	0.041	0.014	0.033	0.063	-	0.005	0.068	0.027	0.019	0.011	0.019	0.019	0.075	-	-	-	0.056	-	0.020	0.017	-	-	0.042
Z	0.002	-	-	-	-	-	-	0.005	-	-	-	-	-	-	-	-	-	-	-	0.034	-	-	0.017	-	-
other	-	0.015	-	-	-	-	-	-	-	-	-	-	-	0.038	0.009	-	-	-	-	-	0.020	-	-	-	0.014

Appendix F - Distribution of European HVS-I haplotypes

The following table shows the polymorphic positions for the 6454 HVS-I sequences and their distribution among the 51 populations studied. Numbers are according to Anderson *et al.*, 1981 (minus 16,000 for brevity). Mutations are transitions unless stated (e.g. 293t indicates a thymine transversion at np 293). (+/d): ins/del, respectively. (pops): number of populations sharing each haplotype (ht), and ($\sum$) total number of individuals sharing each haplotype. Codes for populations are as follows

po: Polonia,
cz: Czech Republic
au: Austria
sz: Switzerland
bv: Bavaria
gs: south Germany
gc: central Germany
ng: north Germany
dn: Denmark
nw: Norway
fn: Finland
sw: Sweden
ic: Iceland
sa: Saami
ka: Karelia
vf: Volga-Finn
es: Estonia
be: Belarus
nr: European Russia
ur: Urlaic region
kr: Ukraine
ws: west Siberia
bl: Bulgaria
rm: Rome
si: Sicily
ni: north Italy (Ladins)
sd: Sardinia
tc: Tuscany
sp: Spain
pt: Portugal
fr: France
bq: Basque Country
gk: Greece
sy: Syria
tk: Turkey
ar: Armenia
ge: Georgia
pl: Palestine
ad: Adygei
gy: Roma-Gypsy
ro: Romania
os: Ossetia
sc: Scotland
ok: Orkney
en: England
cw: Cornwall
ir: Ireland
wa: Wales
ky: Is. Skye
wi: Western Is., and
anc: ancient Britain

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
1	H'	crs	7	11	25	10	8	55	32	3	13	145	39	5	70	1	21	4	29	1	2	3	3	4	28	4	33	7	19	9	44	16	14	31	18	6	5	21	21	12	10	14	13	10	171	14	56	27	26	23	9	26	3	1181	51
2	J'	126	3	3	2	1	3	7	2	2	1	26	9	2	24		2	3	5	1	1	1	1		7		1		1	3	5	1	2		11		3	1	1	4	1	1	1			66	2	17	6	9	7	4	7	3	263	44
3	K	224 311	1	2	5	3	1	4	2	8		18	3	1	19		1	1	2		1	1	1	1	6	1	5	2	2	2	6	3	3	2	3	1		8	7	3	1	2	1	1	31	3	8	1	4	4	2	13		200	46
4	V	298	1	3	1	3	1	1	3	1	1	8	4	13	1	59	2		2	1	1	1			1	1	1	3			6	1	1	5	2			1			2	5		14		7	2	7	2		2		170	37
5	H'	311	3	5	2	9	3		1	6	2	8	3	4	14	1	2	2	2	1		2		3	4		2	2	3		5	1	3	2	5	1	2	2	9	1	1		5	2	16	1	4	3	1	2		2	1	153	44
6	H'	189	2	1	3	1	1	4	3	8		7	6		13		2	2	12	1	2	1	1	1	9	2	1		1		1	2		1	8			1		1		3		16	2	5	3	3		1	2	1	134	38
7	H'	304	5	2		2	1	1		2		14	3		3	2	3		7	1	1	1			3				1	1	3		2	4	4	1	2	2	1	1		2	7	1	13	5	5	1		7		3		117	36
8	T2	126 294 296 304	1		1	1		7	2		3	15	5	1	4		1	3	3		1		2	1	1	1					4	1	1			1				1	1			1	26		4		2				3	98	29
9	T1	126 163 186 189 294	1	3	2		2	2	2	3		4	3	2	2		3		1	1	1		1	9			3			2	2			1	2	2	4	1	2	1	1	4	2	11	2	3	1	2	1			2	91	37
10	H'	129		1	1	2		2				7	1		28				2						1		2	1	2		4			6	1		1			3		1			8	1	7	1			1		1	85	24
11	I	129 223		1	2			3	1	1	1	5	4		9		1			1					1	1									1		1					1	5			2			10	1	73	21
12	U5b1	144 189 270										5	4	4		45	5				1	1																																	65	7
13	U5a1a	256 270		2		2		2		2		4	3		3		4		1		1	1									1				3	1		1		1	2		13			1	2	1	2	4	2	60	25
14	H'	162	1					3	2			6	5	1	2	3	5		7		1	1		2							1						1	3		1		3	2		5	2	2			1		3	2	60	23
15	U5a1	192 256 270		5	2	1		2		1		9	3	1	5	1			1			1			2						3				3						2		10	1	2		1				1	57	21
16	X	189 223 278					1		1			6			1				1			1	1			2	1				1			1	4	1		3		1		6	9	7				1	1	1		50	20
17	J1b1	126 145 172 192 222 261										5			15																																				1		1		45	6
18	H'	261				4	1			1		3		1	1			1	2			1		2	1								2		1						2	16				2		3	1			45	18
19	T	126 294 296		2	2				1	3		3							2	1							3	4						2	1	1	1	2	3	2	2			3	3		1	2				1	45	22
20	U5b	189 270					1	2	1	2		7			6	3	1		2	1		1			2	1	1				1					2			2				2		2		1		1			2	44	22
21	I	129 223 291 298																							1																	42													43	2
22	T	126 294		1			1					3			5								2				1		1									2			1			6	2	3	2	1			8		39	15
23	I	129 172 223 311						1					1		6			1	1			1		1		1				2											11			8		1		1		1	2		39	15
24	W	223 292	1			2	1	2				6	3						1							1		1		2						2			2				6		2	1	3				1	37	17
25	IXW	223	1			1		1		2						1			2	1		2	2	2	3								1	2		1		2		3			2		2					1		32	19
26	H'	263	1	3	1			2				7		1						1														1									11		2								30	10
27	H'	291		1				1				2	4		1						1			2					1			1		3		2		1			1	4		2	2				1		30	17
28	J'	126 261	1	1	1	1		3				2	1										1	2							1					1	1				4	1	3		1			1		3		29	18
29	T2	126 294 304		3	1	1		2	1	1		3																				2		1	1		2			1		3		2		2	1		1		28	17
30	J1a	126 145 231 261		1	3		2	1		2	1	3		1					1										1	1											1	3		6							27	14
31	H'	293 311		2	2				1			1							2						2						2				2						1		9		1	1						26	12
32	U5b1	144 148 189 270												3		23																																					26	2
33	J1b1	126 145 172 222 261										4			7				1			1								1													6		2	1		1					24	9
34	K1	224 311 320										1			11																									1			5	2	1		1			2		24	8
35	T5	126 153 294			1				1			6			12																				1														1	1	23	7
36	H'	162 209						3		1		3							1														1										12		1							22	7
37	U5a	189 192 270					2	1		1		1	8				3														1										2			2					1		22	10
38	U5	270		1				2				4							1									1						1						1			4	3	2		2					22	11
39	H'	249 311		1								2			17														1								1														22	5
40	U5a1	192 256 270 291						1				2	2		2				2			1		1															1				5	1	1			1		1	1	22	14
41	H'	256						1		2	2	5	1		1						1						1										2		1			1	3									21	12
42	H'	221				2	1	1				1			1											1		2		1	1				1						1		1		1	1		3			1	20	16
43	H'	183c 189			3	2						1	2		4		1														2					1								4							20	9
44	J2	126 193											1		5												2				2						2	1	1			2	1				1		2			20	11
45	U5a1	192 256 270 311			1							1							1				1														2		2		1	5		2				1	1	2	20	12
46	T2	126 188 257 294 296													19																																					19	1
47	H'	224						3				1								1	1									1	2		1	1			1					2							5		19	11
48	H'	235 291 293																										1															12		1				1	3		18	5
49	V	153 298								2		3		2		7	2	1	1																																		18	7
50	H'	278															1			1		1									1		1	1	1		1	1					5	1				3				18	12
51	H'	134				1	1		2			2						1		1					2																		5			2						17	9
52	H'	239						2	1			2																			1											2	7			1						17	8
53	J2	126 193 278							1				1		1			1											1		1		1		1								7						2			17	10
54	H'	172		1				1				1			1		1					1							1	3					1		4						1		1							17	12
55	V	261 298 311																																										15				1					16	2
56	H'	129 239													15																																					15	1
57	U5a1	192 256 270 294			1					1	1	1	1						4									1								1			1						1	1						14	11
58	X	126 189a 223 278																																							14													14	1
59	T	126 189				1		1				1					1		1		1						1		1								1							1	1		1				1		13	13
60	H'	179	2		1		1		2						1																	1	2		3																13	8

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
61	K	224 311 319																																										5			1				6		12	3
62	H'	293						1					1	2	3																1	1		1			1															12	9
63	H'	129 316									2		2						1	1					1	1			1																1		2						12	9
64	I	129 223 311				1		1			1				4																					1		1										1			12	9
65	T1	126 186 189 294																																			1								3					8			12	3
66	U2	129c 189						1		1															1		1				1						1							6								12	7
67	H'	192																							1		2					2		2		1	1	1												1		12	9
68	H'	168 192																																					3		7			2										12	3
69	K	129 224 311							1																								1											3		1			1	4		11	6
70	H'	260		1											1												2																	5		1					1		11	6
71	H'	304 305													11																																						11	1
72	H'	167 274 304													11																																						11	1
73	H'	146			1	2						2					1		2																1										1						1		11	8
74	H'	209	1									2																									1		2		1	1				1		1			1			11	9
75	T3	126 292 294							1												1				1				2										2		1			1							2		11	8
76	H'	239g													7																														2				1				10	3
77	U5a1	192 256 270 291 294													10																																						10	1
78	H'	129 256										1			1																														2				1			5	10	5
79	K	224 291 311										1																													1					5						3	10	4
80	H'	278 293 311											2						3	1	1		1																						2								10	6
81	T3	126 292 294 296	1	1								2													1							1				1			1	1		1										10	9
82	H'	278 311										1														1															3		1	4									10	5
83	H'	287				3																										1			4						1									10	5
84	H'	218 328a																							3						1				1		1	2	1	1												10	7
85	J'	126 319							1																																			5		1	1		1				9	5
86	Z	129 185 223 224 260 298										2		1	1	5																																					9	4
87	H'	234	1						1						2																1							1			1		2								9	7
88	J1b	126 145 222 261																			1										2						3		2					1								9	5
89	J'	126 241																								1									8																9	2
90	K	224 245 311			1			1					1								1																								1	2	1							8	7
91	T4	126 294 296 324								2	1						1												1					1												1							8	7
92	I	129 145 223					1					2			1																													3							1		8	5
93	U5a	192 270		1				1				2								1																	1								1		1							8	7
94	H'	274			1				1			2	1												1																						1							8	7
95	H'	292											2															2								1			1		1						1							8	6
96	H'	319																							1																				6						1		8	3
97	H'	213						1	1																													2							1	1	1		1				8	7
98	U1	189 249							1			1															2		1									3															8	5
99	H'	168						1											1						1	1											1	2															8	7
100	W	192 223 292 325											1						2												1								2				1		1							8	6
101	H'	290						1															1														4			1		1										8	5
102	X	189 223 255 278	1										1																		1													1						3		7	5
103	H'	235 291										1																			1			1										2		2							7	5
104	H'	140 293 311																																											3		2			2				7	3
105	H'	172 304 311						1						1																														5									7	3
106	H'	248			2	1			1						1																						2															7	5
107	H'	145		1	1							1																1	1																2									7	6
108	H'	222						3		1	1																														1										1	7	5
109	other	126 136 182c 183c 189 223														7																																					7	1
110	J'	126 278						1																								1		2								1				1		1				7	6
111	H'	176																		1											1	1		2										1					1				7	6
112	H'	299			1										1																						1								2		1				1			7	6
113	V	124 298													5																																					2		7	2
114	H'	129 189						1		1																					1												4									7	4
115	H'	192 304		2	1		1																																2							1							7	5
116	J1b	126 145 222 235 261 271																																							7													7	1
117	J2	126 193 256																													1									1		1		4									7	4
118	J'	126 291						1																								3	1																		6	4
119	H'	189 311								5																						1																					6	2
120	U5a	192 270 319																																6																		6	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
121	H'	242																																			2						4									6	2
122	T	126 189 294 296 298						1		1																																		4									6	3
123	H'	169 241										2																																4									6	2
124	T2	126 278 294 296 304																																										1						2	3		6	3
125	W	223 292 295				1			1				2																		1											1									6	5
126	T	126 189 294 296																	1											1		1										1	2								6	5
127	H'	189 278						1		1							1																		1								2									6	5
128	H'	261 311											1						3															1												1							6	4
129	H'	129 189 249 288																																	1				1										4		6	3
130	H'	295				1			1		1																1																1	1									6	6
131	K	224 234 311		1																												1		1			1			1												6	6
132	V	298 311		1		1																			1								1			1															1	6	6
133	H'	140 265 293 311						1											2						1										2																		6	4
134	U7	318t																							2										2				1				1									6	4
135	H'	301										2																			1												2							1		6	4
136	C	129 223 298 327																																						3			3										6	2
137	H'	304 311													1																						1		1							1		1						5	5
138	U5a	192 270 304										1																																3		1							5	3
139	T2	126 172 294 304			1							1																				1													1		1					5	5
140	H'	172 311			2	2																							1																								5	3
141	J'	126 256										2			1																																					2		5	3
142	K	167 224 311			5																																																5	1
143	H'	169						2																							1				2																	5	3
144	H'	265c							2						2																				1																		5	3
145	H'	148										1																																											5	4
146	U5a1	192 239 256 270													1																																							5	4
147	H'	189 325							2	1					2																																							5	3
148	H'	209 304 305													5																																							5	1
149	J1b1	126 145 172 192 261													3																																								5	3
150	T2	126 147 294 296 297 304																																																				5	1
151	H'	239 311										3																																											5	2
152	J'	126 258c																																																					5	5	1
153	V	126 298										2								1	1																																				5	4
154	H'	261 291 311			1		1																																															5	3
155	H'	169 299																																																					5	2
156	J'	126 311									1	1							1																																					5	5
157	H'	140																		1																																			5	2
158	H'	129 311		1								3																																										5	3
159	U5b	189 270 311		1								1																																										5	5
160	W	223 234 292											2				2				1																																			5	3
161	U7	309 318t							1																																													5	4
162	N	145 176g 223						1																			1		1																									5	5
163	W	172 223 231 292																							1										1		1																	5	4
164	U5a1	192 256 270 304									1	1						1																																						5	4
165	I	129 189 223 249 311																											1		1																								5	4
166	H'	294		1																1			1						1																									5	5
167	I	129 189 223 311																									1												2						1			1						5	4
168	H'	189 192																	1	2																											2									5	3
169	X	189 223 241 278																																																					5	1
170	J1b1	126 145 172 261																	1	1																					2															4	3
171	H'	192 311																	2																							1							1							4	3
172	V	216 298						1																							3																								4	2
173	J'	126 324			1			1	1																1																														4	4
174	H'	207							1																						3																								4	2
175	J'	126 209								1																					1																								4	4
176	T1	126 163 186 189 272 294								1	3																																														4	2
177	U5a1	148 192 256 270										2			2																																									4	2
178	K	222 224 270 311										2																																											4	3
179	H'	293c 311										4																																											4	1
180	IXW	223 278										2																		1																								4	3

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bi	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
181	I	129 223 304										1			2																												1									4	3
182	H'	218			1							1																		1													1									4	4
183	H'	167 294 304																																									4									4	1
184	I	129 172 189 223 311																																									2			2						4	2
185	H'	240										1																		2	1																					4	3
186	H'	266 311			2																															1								1								4	3
187	W	223 270 292																																								1								3	4	2
188	H'	288						1	1																					1							1														4	4
189	W	129 223 292										1					1																			1	1															4	4
190	K	224 278 311																							1	1													1													4	4
191	H'	294 304 320						1	1																																			1								1	4	4
192	J'	126 231										2			1																															1							4	3
193	J'	126 153													1																												1			1	1						4	4
194	H'	220c										1																							1												1			1				4	4
195	H'	249										1			1															1																1							4	4
196	H'	256 311												1											1				1									1															4	4
197	T4	126 294 324										1	1							1																1																	4	4
198	U2	129c 189 256				1																1														1			1													4	4
199	K	189 224 311																																					3								1						4	2
200	H'	188								1																								1	2																	4	3
201	J'	126 222					2			1	1																																											4	3
202	K	172 224 311	1									1														1	1																											4	4
203	J1a	126 145 189 231 261										1								1															1							1											4	4
204	V	189 298																											1					1									1		1							4	4
205	H'	153				1																															2								1						4	3
206	U1	186 189 249											1																1									1									1					4	4
207	IXW	223 291										1										1		2																													4	3
208	T1	126 163 186 189 294 311																																									3		1								4	2
209	J'	126 145																																		1		3														4	2
210	K	224 260 311																																		1		2		1												4	3
211	H'	284																																			1	2					1									4	3
212	U1	129 186 189 249																																						1				3									4	2
213	U5a1	192 217 256 265 270																2	2																																				4	2
214	V	261 298																																										4									4	1
215	H'	193													1																														2								3	2
216	H'	124						1																	1																								1				3	3
217	T5	126 153 294 296																																	1		1															3	3
218	U5a1	172 189 192 270										1																		1																				1				3	3
219	H'	172 256						1			1	1																																									3	3
220	X	183c 189 223 278																															1				1														1	3	3
221	V	240 298										1																		1				1																		3	3
222	V	183 298										1	2																																								3	2
223	V	216 261 298											2																																								3	2
224	K	209 224 311								1					2																																						3	2
225	U5a	189 256 270				1		1																																	1										3	3
226	H'	222a 294 304 311													3																																						3	1
227	H'	129 278 311													3																																						3	1
228	U5a	169a 192 235 270																											3																								3	1
229	H'	257																											3																								3	1
230	U5	270 311										3																																										3	1
231	H'	189 234 325										3																																										3	1
232	H'	269 288																																																				3	1
233	H'	266										1																																										2	2
234	U5a	192 249 270										1																																										2	2
235	H'	231 304 311																																																			3	1
236	H'	153 288																																																			3	1
237	J1a	126 145 183c 189 231 261				3																																															3	1
238	J'	126 176				1						1																	1																								3	3
239	H'	213 304										2																		1																								3	2
240	U5a1	168 192 256 270			1									1																						1																	3	3

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bi	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pi	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
241	T2	126 192 294 304										1																																						2		3	2
242	T	126 192 294																													1	1																		1		3	3
243	U5	224 242 270																																																3		3	1
244	X	189 223 278 293																							2															1														3	2
245	H'	207 304																																				1									2						3	2
246	H'	257 298											3																																								3	1
247	U5a1a	256 270 294										1	1																																			1						3	3
248	H'	189 290					1	1															1																														3	3
249	T2 [1]	126 304						1													1																												1					3	3
250	U5	261 270													2								1																															3	2
251	T2 [2]	294 296 304																				1						1						1																				3	3
252	T3	126 189 292 294																1											2																									3	2
253	X	189 223 265 278										1															1																											3	3
254	J'	126 221		1								2																																										3	2
255	H'	162 189										1				1						1																																	3	3
256	T1 [1]	126 163 186 189																1																															2					3	2
257	H'	311 317t																																															1		2				3	2
258	H'	271																					1																												1		1			3	3
259	H'	186 189																				1									2																							3	2
260	I	129 223 264 270 311 319																													2									1														3	2
261	T1	126 163 189 243 294																																			2			1														3	2
262	H'	189 304			1																						1											1															3	3
263	H'	193 249																								1						1						1														3	3
264	H'	261 304																		1												1									1												3	3
265	H'	320								1																																					2						3	2
266	H'	167																			1																												1	1					3	3
267	H'	189 234 324						1																													2																3	2
268	T1 [2]	163 186 189 294							1												1		1																																3	3
269	J'	126 224																																															3						3	1
270	H'	214										1																1																						1				3	3
271	J1b	126 145 222 261 274																											1									2															3	2
272	H'	244																							1			1											1														3	3
273	U5a1	192 256 270 271 294														1	2																																						3	2
274	IXW	223 245															1					1																	1															3	3
275	J'	126 148																																															2				1			3	2
276	U5a1a	256 270 311											1										1					1																											3	3
277	H'	312																																															3						3	1
278	H'	129 221																										2								1																3	2
279	I	129 223 230 233 304																																																3						3	1
280	I	129 223 287 311																																		3																		3	1
281	K	167 209 224 311 319																																				3																3	1
282	IXW	172 223 274 291																																															3						3	1
283	H'	189 293 311																																							1	2														3	2
284	H'	189 295																																				3																3	1
285	IXW	201 223 265																																				2			1														3	2
286	H'	220c 234g 278 292																																				3																3	1
287	K	224 261 311																			2																																	3	2
288	H'	193 219																																															2						2	1
289	U2	129c 183c 189											2																																										2	1
290	U5b	183c 189 270				1							1																																										2	2
291	U5a1a	256 270 291										1	1																																										2	2
292	H'	294 304		1																																													1						2	2
293	H'	162 311																			1																																			2	2
294	H'	183							1			1																																										2	2
295	U5a1	129 192 256 270 291								1		1																																											2	2
296	H'	221 266										1																																											2	2
297	C	223 298 311 325 327													2																																							2	1
298	H'	259						1				1																																										2	2
299	U5	270 292										1																			1																							2	2
300	H'	260 304			2																																																	2	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
301	J'	126 201 265 319			2																																															2	1
302	H'	221 291																															2																			2	1
303	H'	136 295																																														2					2	1
304	u5a1a	129 256 270 286 320									2																																											2	1
305	T2	126 246 248a 294 296 304											2																																										2	1
306	K	192 224 311											1																																				1					2	2
307	U5b	189 270 319											2																																									2	1
308	IXW	189 223 291a											2																																										2	1
309	IXW	223 288											2																																										2	1
310	V	169 298											2																																										2	1
311	H'	147										1	1																																										2	2
312	H'	145 222						1																							1																							2	2
313	H'	192 214a						1																																															2	2
314	H'	205g						2																																															2	1
315	U5a1	192 256 270 286 320							1		1																																													2	2
316	H'	256 304							1																																													2	2
317	H'	129 134										1					1																																						2	2
318	H'	129 304																	1																																				2	2
319	H'	274 304							1						1																																							2	2
320	H'	179 291					1		1																																													2	2
321	H'	184																													1								1														2	2
322	H'	188g																							1																														2	2
323	IXW	189 223																													1										1												2	2
324	H'	162 293 311								2																																												2	1
325	H'	215 263a						1							1																																							2	2
326	K	162 224 311						1																																														2	2
327	H'	265																																				1								1							2	2
328	H'	165 222								2																																												2	1
329	H'	325																					1																																2	2
330	V	153 298 311								2																																												2	1
331	H'	176 311								2																																												2	1
332	U5a	192 235 270 304													2																																						2	1
333	J'	126 178													2																																						2	1
334	J'	126 187													2																																						2	1
335	H'	239a													2																																						2	1
336	U2	129c 183c 209 260			1																							1																									2	2
337	T3	126 145 292 294 296																										2																									2	1
338	U5a1a	187 256 270																																																		2	2	1
339	H'	184 304										1																					1																			2	2
340	J'	126 153 319										2																																										2	1
341	U5a1	145 189 192 256 270 299										2																																										2	1
342	K	189 224 293 311										2																																									2	1
343	H'	189 271										2																																									2	1
344	H'	162 266										1																																									2	2
345	V	271 298										2																																									2	1
346	H'	295g										2																																									2	1
347	H'	186										2																																									2	1
348	H'	235 291 292 319																																														2					2	1
349	J'	126 325			1																																											1					2	2
350	H'	234 311																																			1												1					2	2
351	U6a	172 189 219 278																															2																			2	1
352	V	204 213t 255t 298												1		1																																						2	2
353	H'	152												1		1																																						2	2
354	U5b1	144 189 248g 270												1		1																																						2	2
355	H'	189 300																													1																						2	2
356	U5	193 270 296																																																			2	1
357	J'	126 176 261 266																																																			2	1
358	U5b	189 249 270																																																			2	1
359	J'	126 243										1																																										2	2
360	H'	172 304																																																			2	2

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
361	T	126 256 294 296										1																															1									2	2
362	H'	309																																										1					1				2	2
363	T	126 291 294 296													1																												1									2	2
364	T2	126 292 294 304																																									2									2	1
365	J1a	126 145 231 261 311																																		1							1									2	2
366	H'	209 260 298 t																																									2									2	1
367	U5a1a	256 270 286 298																																									2									2	1
368	U5a1	189 192 256 270 291									1																																1									2	2
369	H'	140 311																																									2									2	1
370	J2	126 193 217																																									2									2	1
371	T2	126 187 294 296 304																																									2									2	1
372	H'	187				1						1																																							2	2
373	K	223 224 311 319				2																																												2	1
374	H'	189 215 263a				2																																												2	1
375	W	223 292 311																													2																				2	1
376	H'	305																													1																			1		2	2
377	U5a1	189 192 256 270										1																			1																						2	2
378	H'	235 291 311												1																	1																						2	2
379	J2	126 193 311																												1							1															2	2
380	H'	129 220c																																1													1					2	2
381	I	129 223 235 260																																													2						2	1
382	I	129 223 293c 296i 311																																															2				2	1
383	H'	172 240 261																																															2				2	1
384	H'	132																																															2				2	1
385	H'	189 207																																																	2		2	1
386	J'	126 242																																														1				1		2	2
387	J1b1	126 145 172 189 192 222 261																																														1				1		2	2
388	K	223 224 234 311																																																	1		1		2	2
389	H'	327			1													1																																				2	2
390	U5a1	129 192 256 270			1																					1																											2	2
391	I	129 153 214g 223 271			1															1																																	2	2
392	U5a1	145 189 192 256 270			1															1																																	2	2
393	K	174 224 311															1																			1																2	2
394	T2 [1]	126 298 304									1																															1										2	2
395	U5b1	144 189 270 301											2																																							2	1
396	H'	271 311																							1								1																			2	2
397	K	224 301 311						1																		1																											2	2
398	H'	189 294								1										1																																	2	2
399	J1a	126 145 231						1														1																															2	2
400	I	129 223 279 311							2																																												2	1
401	H'	214a 249 327						1																														1														2	2
402	W	223 292 295 324																	1																												1							2	2
403	J2	126 193 256 300 309																												1																							2	2
404	J'	126 230																												1																							2	2
405	H'	309 318c																												1																							2	2
406	T1	126 163 184 186 189 294																											2																								2	1
407	X	189 223 278 297																									1			1																							2	2
408	other	126 270										1										1																															2	2
409	H'	262										1															1																										2	2
410	H'	235 261 291										1								1																																	2	2
411	U5b1 [3]	144 148 189														2																																					2	1
412	K	224 293 311																																																		1		2	2
413	K1	189 224 311 320																																																		2		2	1
414	T4	126 187 294 296 324																																																		1		2	2
415	H'	263 278																			1																																	1		2	2
416	X	189 192 223 278																																																		1		2	2
417	B?	172 187 189 217 223																																																		1		2	2
418	T2	126 172 294 296 304																																																		1		2	2
419	U5	270 291																												1																							2	2
420	X	189 223 248 278																																				1														2	2

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
421	H'	187 189																												1										1												2	2
422	U1	129 189 249																																	1				1												2	2
423	IXW	223 311																																		1			1												2	2
424	K	224 304 311																	1																	1																2	2
425	K	224 246t 311																																	1									1								2	2
426	U5	239 270																																									1				1				2	2
427	T2	126 193 294 296 304															1																										1								2	2
428	T5	126 153 207t 294																		1																							1								2	2
429	H'	212																																						1			1								2	2
430	IXW	223 278 294 309																																							1			1								2	2
431	J'	126 265 319																							1																							1				2	2
432	H'	162 292																																																2	2	1
433	V	218 239a 298																									1																							1	2	2
434	H'	289																										1											1												2	2
435	K	124 224 311																																						1				1								2	2
436	J1b	126 145 218 222 261		1																																			1													2	2
437	J1	126 145 261																																					1		1											2	2
438	T1	126 163 172 186 189 294																																					1	1												2	2
439	T1	126 163 186 189 262 294																																					2													2	1
440	T	126 167 186 189 294																									2																								2	1
441	J2	126 193 300																																					1		1											2	2
442	J'	126 258 319																																										2								2	1
443	T	126 294 295																																									2								2	1
444	U1	129 189 220c 249 288																									2																								2	1
445	U2	129c																																											2								2	1
446	N	134 145 176g 223																																					2													2	1
447	H'	134 153 172															1		1																																		2	2
448	H'	134 310 311																																					2													2	1
449	U5b1	144 189 270 291																																					2													2	1
450	IXW	145 176 223																																						2												2	1
451	U5b	147 189 270																								1																										2	2
452	H'	148 256 319																							1															1												2	2
453	IXW	168 201 223 265																																					2													2	1
454	H'	179 227 245 266 274 278																																					1	1												2	2
455	H'	183c 189 234 294																																						2												2	1
456	U5a1	189 192 256 270 294																	1	1																																		2	2
457	U5a	189 203 256 270																																					2													2	1
458	H'	189 222																																					1							1						2	2
459	X	189 223 255 278 300		2																																																2	1
460	X	189 223 278 311																																					1							1						2	2
461	W	192 223 292 304 325																																											2						2	1
462	U5a	192 249 270 311										1																																									2	2
463	H'	207 309 318c																																					1		1												2	2
464	H'	214 244																																						2													2	1
465	H'	214a 231 249 327																																											2								2	1
466	H'	218 297																																											2								2	1
467	K	224 261 293 311										2																																										2	1
468	H'	242 261																																							2													2	1
469	H'	249 311 327																																					2													2	1
470	H'	256 294 304 320																									2																									2	1
471	J'	126 300																																										1							1			2	2
472	W	172 223 292 294																																										1	1									2	2
473	W	184 223 292																	1																																				1	2	2
474	T	126 182c 183c 189 294 296 298			1																																																	1	1
475	T2	126 261 294 296 304							1																																													1	1
476	U5a1	192 256 270 320																																																1				1	1
477	H'	209 311																															1																				1	1
478	H'	153 304		1																																																		1	1
479	H'	162 172 209																																																			1	1
480	J1b1	126 145 172 222																																																		1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
481	U5a1	192 222 256 270																											1																							1	1
482	J1	126 261 311										1																																									1	1
483	J1a [4]	145 231 261																																																			1	1
484	N	129 147a 154 172 223 248 320										1																																									1	1
485	H'	221 230						1																																													1	1
486	T	126 182c 183c 189 294 296																																						1													1	1
487	T	126 209 294 296																																				1															1	1
488	T2	126 266 294 304																												1																							1	1
489	I	129 172 223 311 319		1																																																	1	1
490	H'	134 234												1																																								1	1
491	H'	170			1																																																	1	1
492	U5a1	174 189 192 270 311																																																				1	1
493	H'	189 299																																																				1	1
494	U2	129c 182c 183c 189																																1																				1	1
495	IXW	201 223 256 265										1																																											1	1
496	U5a1a	223 256 270							1																																														1	1
497	J'	126 163							1																																														1	1
498	J2	126 193 274 278																																																			1	1
499	J2	126 193 278 311																										1																									1	1
500	H'	233 256 311			1																																																1	1
501	H'	274 311			1																																																1	1
502	H'	213 263 311			1																																																1	1
503	H'	257 265			1																																																1	1
504	H'	189 261			1																																																1	1
505	H'	134 189			1																																																1	1
506	W	184 223 292 295			1																																															1	1
507	J'	126 250 261 292a			1																																															1	1
508	J'	126 182c 183c 189 249			1																																															1	1
509	J'	126 150 222			1																																															1	1
510	J'	126 169 264			1																																															1	1
511	X	145 179 183c 189 223 278			1																																															1	1
512	K	189a 224 274 311			1																																															1	1
513	H'	304 319			1																																															1	1
514	T	126 185 189 294 296			1																																															1	1
515	I	129 148 192 223 294			1																																															1	1
516	H'	183c 189 265			1																																															1	1
517	U5a1a	256 270 301			1																																															1	1
518	U5a	192 270 294																																1																			1	1
519	U5a1	187 192 270																																1																			1	1
520	H'	215																																1																			1	1
521	H'	235t 286																																1																			1	1
522	T3	126 292 294 311																																1																			1	1
523	K	145 224 311																																1																			1	1
524	U5a	167 192 270 311																																1																			1	1
525	U5a	192 270 311																																1																			1	1
526	H'	189 286 299																																1																			1	1
527	H'	320 323 325																																1																			1	1
528	T2	126 129 165 294 296 304																							1																													1	1
529	H'	129 134 278																							1																													1	1
530	H'	174 192 311																							1																													1	1
531	K	195 224 311																							1																													1	1
532	U5a1	136 192 215 256 270																							1																													1	1
533	H'	192 256 302 320																							1																													1	1
534	U5b	184++189 205 270																																																			1	1	1
535	T1	126 163 186 189 269c 294																																																				1	1	1
536	J'	126 265c 319																																																			1	1	1
537	I	129 223 318																																																			1	1	1
538	J'	126 297																																																			1	1	1
539	J'	126 186 192																																																			1	1	1
540	H'	186 193 219 260																																																			1	1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
541	H'	129 299																																													1						1	1
542	T	126 294 301																																														1						1	1
543	H'	192 218 311 320																																														1						1	1
544	H'	214 311																																														1						1	1
545	H'	189 221 234 290 324																																														1						1	1
546	U5a1	189 192 270 311 336																																														1						1	1
547	J'	126 145 184																																														1						1	1
548	H'	192 294 304 320									1																																												1	1
549	W	223 292 295 304									1																																												1	1
550	K	222 224 311									1																																												1	1
551	J1a	126 145 231 261 272 294									1																																												1	1
552	U5	129 270									1																																												1	1
553	J1a	126 145 231 261 320									1																																												1	1
554	H'	215t 232											1																																										1	1
555	other	126 223 260 298											1																																										1	1
556	H'	258 311 325											1																																										1	1
557	H'	166 304											1																																										1	1
558	H'	217 304											1																																										1	1
559	IXW	182c 183c 183i 189 223 255 278											1																																										1	1
560	U5a1a	162 256 270											1																																										1	1
561	W	223 242 292											1																																										1	1
562	U5b	189 228g 270 317											1																																										1	1
563	IXW	183c 183i 189 223																															1																				1	1
564	V	172 180 298 311																															1																				1	1
565	H'	224 265																															1																				1	1
566	H'	140 183c 189 266a																															1																				1	1
567	U5a1	189 192 270 286																															1																				1	1
568	H'	296																															1																				1	1
569	C	189 223 298 319 327																															1																				1	1
570	J1b1	126 145 146 172 175c 222 261																															1																				1	1
571	H'	162 316																															1																				1	1
572	IXW	223 234 248 265c																															1																				1	1
573	H'	224 316																															1																				1	1
574	J2	124 126 193 291																															1																				1	1
575	V	221 294 296 298								1																																												1	1
576	T3	126 292 294 296 298								1																																												1	1
577	K	214a 224 311								1																																												1	1
578	T	126 140 189 294 296 311						1																																													1	1
579	H'	278 293						1																																													1	1
580	H'	201						1																																													1	1
581	H'	183c 189 293						1																																													1	1
582	H'	183t 188 189						1																																													1	1
583	H'	150 224						1																																													1	1
584	L1b	126 175 189 223 264 270 278 311								1																																												1	1
585	H'	239 286						1																																													1	1
586	T1	126 163 172 186 189 294 298						1																																													1	1
587	H'	188a						1																																													1	1
588	H'	168 311																																						1													1	1
589	J'	126 172						1																																													1	1
590	J1a	126 145 231 261 264 270 278 311								1																																												1	1
591	W	173 223 292						1																																													1	1
592	H'	192a						1																																													1	1
593	H'	134 269 311						1																																													1	1
594	K	224 233 311							1																																												1	1
595	H'	210 304								1																																											1	1
596	U5a	192 257 270							1																																												1	1
597	H'	240c																			1																																	1	1
598	U5a1a	256 270 290 311 320						1																																													1	1
599	U5b	189 263 270							1																																												1	1
600	H'	255 278 294						1																																												1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
601	J1a	126 145 231 261 294								1																																											1	1
602	K1	224 311 325						1																																													1	1
603	IXW	189 223 255 279							1																																												1	1
604	T2	126 178 294 296 304																	1																																			1	1
605	IXW	189 223 279							1																																													1	1
606	T	126 183c 189 294 296																																					1														1	1
607	H'	215 304								1																																												1	1
608	T2	126 188 189 280 294 296 304							1																																													1	1
609	T2	126 234 294 296 304																																					1														1	1
610	V	126 298 311							1																																												1	1
611	H'	162 248 266 311								1																																												1	1
612	T1	126 163 186 189 267 294								1																																												1	1
613	J1	126 145 261 278							1																																												1	1
614	V	291 298								1																																													1	1
615	L1a	129 148 168 172 187 188g 189 223 230 311 320																																							1												1	1
616	H'	146 154 186 189 267 294								1																																												1	1
617	H'	221 266 316							1																																												1	1
618	H'	129 154 186 189 267 294								1																																												1	1
619	J1b1	126 145 172 261 292						1																																													1	1
620	U5a1	192 256 270 293							1																																												1	1
621	U5b	174 189 270 311						1																																												1	1
622	H'	183i 189 217 260							1																																											1	1
623	V	224 298								1																																											1	1
624	H'	172 207 209								1																																										1	1
625	N1b	126 145 176a								1																																										1	1
626	H'	302							1																																											1	1
627	N1b	129 150a 176g 223						1																																											1	1
628	IXW	158 223 271							1																																											1	1
629	U2	129c 183i 184i 189							1																																											1	1
630	U1	189 192 249							1																																											1	1
631	V	298 325								1																																										1	1
632	K	162 224 288 311						1																																											1	1
633	IXW	223 294								1																																										1	1
634	U5b	128g 189 270 319						1																																											1	1
635	H'	327a							1																																											1	1
636	H'	184 311																																		1																1	1
637	H'	185 189																						1																												1	1
638	H'	203 293 311								1																																										1	1
639	H'	263 311						1																																											1	1
640	U5a1a	256 270 278						1																																											1	1
641	H'	207 255							1																																											1	1
642	H'	231 311 319						1																																											1	1
643	T2	126 189 234 294 296 304							1																																											1	1
644	J1a	126 145 231 261 299																																					1														1	1
645	U5a1a	193 256 270 292						1																																											1	1
646	K	224 296 311						1																																											1	1
647	H'	209 304																			1																															1	1
648	IXW	223 286																																																	1			1	1
649	H'	189 263 304 320								1																																											1	1
650	H'	162 303						1																																												1	1
651	U5a1	183i 184i 189 191 192 270						1																																												1	1
652	H'	234 292 324						1																																												1	1
653	T1 [1]	126 163 186 189 298								1																																											1	1
654	H'	129 294								1																																											1	1
655	H'	192 249						1																																												1	1
656	H'	264																																					1														1	1
657	W	223 292 294								1																																											1	1
658	H'	263 316						1																																												1	1
659	H'	224 263						1																																												1	1
660	H'	300																																							1												1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pi	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
661	H'	189 294 304						1																																												1	1
662	H'	148 214							1																																												1	1
663	U5a1	145 187 189 192 256 270 311								1																																											1	1
664	J'	126 327								1																																											1	1
665	J'	126 248																																							1											1	1
666	H'	188 189 290 319						1																																												1	1
667	IXW	223 249 286							1																																												1	1
668	U5a1a	146 256 270								1																																											1	1
669	H'	189 234													1																																						1	1
670	H'	328													1																																						1	1
671	J'	126 168 178													1																																						1	1
672	J'	126 182c 183c 189 290													1																																						1	1
673	H'	218 304 305													1																																						1	1
674	H'	147 234 304													1																																						1	1
675	V	124 294a 298													1																																						1	1
676	T2	126 188 192 257 294 296													1																																						1	1
677	T2	126 188 232 257 294 296													1																																						1	1
678	T	126 232 257 294 296													1																																						1	1
679	T2	126 188 294 296													1																																						1	1
680	K	126 145 224 311													1																																						1	1
681	K	224 274 311													1																																						1	1
682	J1b1	126 145 172 192 261 271													1																																						1	1
683	U5a1a	256 270 291 294													1																																						1	1
684	X	183i 189 223 278													1																																						1	1
685	H'	183c 183i 189													1																																						1	1
686	J1a	126 145 184i 189 231 261													1																																						1	1
687	H'	294 304 311													1																																						1	1
688	IXW	209 223													1																																						1	1
689	H'	129 167 274 304													1																																						1	1
690	K	207c 224 311													1																																						1	1
691	H'	248 283													1																																						1	1
692	H'	213 248 311													1																																						1	1
693	IXW	192 223 325													1																																						1	1
694	H'	129 274													1																																						1	1
695	H'	189 278 293 311																										1																									1	1
696	H'	153 294 296																											1																							1	1
697	K	129 224 240t 311																										1																									1	1
698	T2	126 265 294 296 304																											1																							1	1
699	H'	248 260																									1																									1	1
700	T	126 184 221 280 294 296 308																									1																									1	1
701	T	126 184 280 294 296																									1																									1	1
702	J'	126 260 308																									1																									1	1
703	T	126 186 189 294 296 299 315																									1																									1	1
704	H'	183c 209 252 260 301 313																									1																									1	1
705	H'	163 186 211																									1																									1	1
706	T1	126 163 186 189 246 294 299																									1																									1	1
707	H'	182c 183c 189 221 260																									1																									1	1
708	T2 [1]	126 278 296 304																									1																									1	1
709	J'	126 163 260 313																									1																									1	1
710	H'	134 263																											1																							1	1
711	X	189 192 223 278 294																										1																									1	1
712	W	192 223 291 292 325																											1																							1	1
713	H'	189 284 286																											1																							1	1
714	H'	203 221																									1																									1	1
715	H'	145 214																									1																									1	1
716	H'	162 278 304																										1																									1	1
717	T	126 280 294																									1																									1	1
718	H'	172 213																											1																							1	1
719	H'	235 319 320																										1																									1	1
720	L1a	129 148 168 172 187 189 223 230 311 320																										1																									1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
721	U1	145 188g 189 249																											1																								1	1
722	U5	270 296 325																											1																								1	1
723	K	180 224 249 311																											1																								1	1
724	U5a	169a 192 270																											1																								1	1
725	H'	291 324																											1																								1	1
726	T5	126 153 266 294 304																											1																								1	1
727	T1 [42]	163 186 189 274 294																												1																								1	1
728	K	156 224 240c 311																												1																								1	1
729	T	126 145 184 280 294 296																										1																										1	1
730	H'	189 231 272																										1																										1	1
731	H'	134 189 221 278 313																										1																										1	1
732	U5a	192 270 290																												1																								1	1
733	J'	126 296																										1																										1	1
734	T	126 247 294 296																										1																										1	1
735	T	126 275 294 296 326																										1																										1	1
736	T	126 244 294 296																										1																										1	1
737	H'	124 221 246 301 311																										1																										1	1
738	H'	192 261 311																												1																								1	1
739	U7	129 189 318t																											1																										1	1
740	X	189 223 248 278 324																												1																								1	1
741	T1	126 163 186 189 221 294																										1																										1	1
742	U5b	189 221 222 270																										1																										1	1
743	H'	222 278 294																												1																								1	1
744	U5a1a	246 256 270																										1																										1	1
745	U5b	186 189 270																										1																										1	1
746	H'	284 286 292																										1																										1	1
747	K	221 224 311																										1																										1	1
748	U5	227 270 311																										1																										1	1
749	H'	221 224																										1																										1	1
750	H'	224 233																										1																										1	1
751	K	224 256 311																										1																										1	1
752	H'	147 189																												1																								1	1
753	H'	220c 290																																																					1	1	1
754	J'	126 163 266 311																																																					1	1	1
755	U5b	183 189 270										1																																												1	1
756	H'	172 301										1																																												1	1
757	H'	172 256 304 311										1																																												1	1
758	H'	180 311										1																																												1	1
759	H'	293 304 311										1																																												1	1
760	U5a1	182 192 270 311										1																																												1	1
761	V	234 298										1																																												1	1
762	IXW	192 223 242 270 304										1																																												1	1
763	W	223 234 245 292										1																																												1	1
764	H'	256 279										1																																												1	1
765	H'	299 311										1																																												1	1
766	H'	206 215 265										1																																												1	1
767	H'	215 265										1																																												1	1
768	H'	170c 189										1																																												1	1
769	IXW	178 183d 189 223										1																																												1	1
770	X	183d 189 223 278										1																																												1	1
771	H'	183d 189 274										1																																												1	1
772	K	183d 215 224 311										1																																												1	1
773	K	224 242g 291 311										1																																												1	1
774	U5a1	192 256 266 270 291										1																																												1	1
775	H'	129 134 311										1																																												1	1
776	N1b	129 145 154g 176g 223 256 291										1																																												1	1
777	I	129 223 224 260 298										1																																												1	1
778	J'	126 226t										1																																												1	1
779	J'	126 189 265 319										1																																												1	1
780	T2	126 227 294 296 304										1																																												1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
781	T2	126 291 294 296 304										1																																									1	1
782	T3	126 192 266g 294 296 324										1																																									1	1
783	U5b1	144 184d 185d 189 194c 270										1																																									1	1
784	T1	126 132 182d 183d 189 294 296 298										1																																									1	1
785	T	126 182d 183d 189 294 296 298										1																																									1	1
786	J1a	126 145 231 261 262										1																																									1	1
787	T2	126 293 294 296 304										1																																									1	1
788	T3	126 278 292 294 296 304										1																																									1	1
789	A	126 223 235 290 319										1																																									1	1
790	J'	126 145 195										1																																									1	1
791	J'	126 187 298										1																																									1	1
792	U5a1	179 192 256 270 294										1																																									1	1
793	U5	147a 167 192 270										1																																									1	1
794	U5a1	192 256 270 295g										1																																									1	1
795	U5	144 270										1																																									1	1
796	U5	270 312t										1																																									1	1
797	I	129 223 271g 293c 311										1																																									1	1
798	K1	224 290 311 320										1																																									1	1
799	K	144 224 311										1																																									1	1
800	H'	188 286g 311										1																																									1	1
801	K	215 224 311										1																																									1	1
802	K1	214g 224 311 320										1																																									1	1
803	H'	281c 293 311										1																																									1	1
804	V	260 298										1																																									1	1
805	Z	129 185 223 224 298										1																																									1	1
806	U2	129c 183c 209										1																																									1	1
807	N1b	129 145 176g 192 223 256 304										1																																									1	1
808	H'	242 304										1																																									1	1
809	H'	295g 304										1																																									1	1
810	H'	162 209 235										1																																									1	1
811	W	223 278 292 295										1																																									1	1
812	V	171 180 298										1																																									1	1
813	N	145 154g 176g 223 256 291										1																																									1	1
814	H'	261 284										1																																									1	1
815	H'	125c										1																																									1	1
816	H'	144										1																																									1	1
817	H'	193 318										1																																									1	1
818	H'	218 278										1																																									1	1
819	H'	218 295g										1																																									1	1
820	H'	241										1																																									1	1
821	H'	261 275										1																																									1	1
822	H'	288 293										1																																									1	1
823	W	163c 223 292										1																																									1	1
824	H'	129 171										1																																									1	1
825	H'	219										1																																									1	1
826	H'	224 318										1																																									1	1
827	H'	254 304										1																																									1	1
828	H'	317										1																																									1	1
829	H'	195 291										1																																									1	1
830	H'	147 303										1																																									1	1
831	H'	235 263 291																																															1					1	1
832	U5b	189 193d 249 270																																															1					1	1
833	J'	126 220c 258c																																															1					1	1
834	T1	126 163 186 193d 294																																															1					1	1
835	IXW	223 278 318																																															1					1	1
836	U1	189 249 311																																															1					1	1
837	H'	180 219 221																													1																								1	1
838	V	242 298 311																													1																								1	1
839	H'	304 327																													1																								1	1
840	IXW	209 223 311																													1																								1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
841	X	189 192 223 260 278 294																													1																							1	1
842	U6b	163 164t 172 219 311																													1																							1	1
843	J'	126 264																													1																							1	1
844	H'	153 162																													1																							1	1
845	H'	172 210																													1																							1	1
846	other	126 182c 183c 189 223														1																																							1	1
847	U5b [3]	144 189														1																																							1	1
848	V	144 298														1																																							1	1
849	V	144 189 298														1																																							1	1
850	other	126 136 189 223														1																																							1	1
851	U5b1	144 189 270 302c														1																																							1	1
852	U5b1	144 189 270 302														1																																							1	1
853	U5b1	144 175 189 270														1																																							1	1
854	V	153 182 298														1																																							1	1
855	H'	162 233t														1																																							1	1
856	Z	129 185 223 224 260														1																																							1	1
857	X	189 223 278 287																																																				1	1	1
858	T	126 189 294 298																																																			1	1	1
859	W	189 223 292 295 324																																																			1	1	1
860	T2	126 172 294 304 311																																																			1	1	1
861	V	298 301																																																			1	1	1
862	H'	235 291 297																																																			1	1	1
863	T2 [1]	126 189 296 304																																																			1	1	1
864	U2	128g 129c 189																																																			1	1	1
865	T	126 257 294																																																			1	1	1
866	J'	126 271																																																			1	1	1
867	H'	169 256 311																																																			1	1	1
868	H'	269 304																																																			1	1	1
869	T4	126 294 296 300 324																																																			1	1	1
870	U5a1a	256 262 270 278																																																			1	1	1
871	T2	126 239 294 304																																																			1	1	1
872	H'	254																																																			1	1	1
873	J1b1 [2]	145 172 192 222 261																																																			1	1	1
874	H'	189 239																																																			1	1	1
875	K	168 224 311																																																			1	1	1
876	H'	240 309																																																			1	1	1
877	T1	126 163 186 189 263 294																																																			1	1	1
878	T1	126 163 186 189 257 294																																																			1	1	1
879	V	214 298																																																			1	1	1
880	H'	146 260																																																			1	1	1
881	IXW	172 223 311																																																			1	1	1
882	J'	126 235 311																																																			1	1	1
883	T2	126 189 294 304																																																			1	1	1
884	U5	270 296																																																			1	1	1
885	T3	126 145 243 292 294																																																			1	1	1
886	IXW	162 209 223 278																																																			1	1	1
887	J'	126 295 299t																																																			1	1	1
888	H'	192 219																																																			1	1	1
889	H'	235 291 292																																																			1	1	1
890	K	224 266 311																																																			1	1	1
891	J2	126 193 248 249																																																			1	1	1
892	U5a1a	256 270 293 311																																																			1	1	1
893	H'	189 293																																																			1	1	1
894	H'	179 234																																																			1	1	1
895	other	129 172 188a 189 298																																																			1	1	1
896	T1	126 163 186 189 269 294																																																			1	1	1
897	U5b	145 189 270 299																																																			1	1	1
898	IXW	129 189 223 249 299 311																																																			1	1	1
899	I	129 214 223																																																			1	1	1
900	H'	224g																																																			1	1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
901	U5	162 270																																										1									1	1
902	H'	162d																																											1									1	1
903	T2	126 189 248g 294 296 298																																											1									1	1
904	T2	126 135t 294 296 304																																											1									1	1
905	K	189 224 250g 252 311 321																																											1									1	1
906	I	129 145 189 223																																											1									1	1
907	H'	167 299																																											1									1	1
908	H'	261 278																																											1									1	1
909	H'	162 184 209																																											1									1	1
910	T	126 189 291 294 296																																											1									1	1
911	H'	174 189 311																																											1									1	1
912	U5b	189 235 269c 270																																											1									1	1
913	H'	183 265c																																											1									1	1
914	IXW	183 223 278																																											1									1	1
915	H'	162 209 278																																											1									1	1
916	H'	167 179																																											1									1	1
917	H'	129 162																																											1									1	1
918	T5	126 129 153 294																																											1									1	1
919	U5a1a	172 256 270																																											1									1	1
920	V	278 298 301																																											1									1	1
921	U5a	167 192 270																																											1									1	1
922	J1b1	126 145 172 222 261 287																																												1									1	1
923	U5b	183d 189 270																																												1									1	1
924	U5a1a	256 270 292																																												1									1	1
925	K	192 222 224 270 311																																												1									1	1
926	H'	293 294 311				1																																																	1	1
927	U5	270 290				1																																																	1	1
928	H'	176 219				1																																																	1	1
929	V	183c 189 298				1																																																	1	1
930	H'	295 304				1																																																	1	1
931	H'	217 218																												1																									1	1
932	K	224 311 318t 319																												1																									1	1
933	IXW	124 223 278																												1																									1	1
934	H'	241 287																												1																									1	1
935	U6b	129 163 164t 172 219 311																												1																									1	1
936	H'	239 242																												1																									1	1
937	IXW	129 213 223 234 278																												1																									1	1
938	IXW	169 193 195 223 243 261																												1																									1	1
939	V	249 298																												1																									1	1
940	W	223 292 320																												1																									1	1
941	H'	207 289																												1																									1	1
942	H'	221 328a																												1																									1	1
943	B?	172 189 223 278																												1																									1	1
944	H'	311 319 321																												1																									1	1
945	H'	125 181t 189 295																												1																									1	1
946	H'	192 228 319 321																												1																									1	1
947	H'	124 209 278																												1																									1	1
948	U5a	192 251 259 270 304																												1																									1	1
949	T3	126 243 278 292 294 296																												1																									1	1
950	H'	134 304																												1																									1	1
951	H'	189 266																												1																									1	1
952	H'	292 293 311																												1																									1	1
953	T	126 172 245 294 296																												1																									1	1
954	IXW	124 223 234 278																												1																									1	1
955	H'	129t 189																												1																									1	1
956	U6a	172 219 235 278																												1																									1	1
957	H'	173 290																																		1																		1	1
958	U2	129c 189 277																																		1																		1	1
959	H'	189 261 262																																		1																		1	1
960	H'	163 168 227 285																																		1																		1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
961	T1	126 163 186 189 287 291 294																																	1																		1	1
962	H'	134 176																																	1																		1	1
963	H'	168 212 239																																	1																		1	1
964	H'	173 261 290																																	1																		1	1
965	W	223 286 292																																	1																		1	1
966	N1b	126 145 176g 223 311																																	1																		1	1
967	H'	266 287 318t																																	1																		1	1
968	K	189 222 224 311																																	1																		1	1
969	H'	235 304																																	1																		1	1
970	H'	278 318c																																	1																		1	1
971	H'	278 325																																	1																		1	1
972	H'	172 222 256																																	1																		1	1
973	U7	129 318t												1																																									1	1
974	T2	126 172 294 299 304												1																																									1	1
975	H'	169 265c 284 304												1																																									1	1
976	H'	134 240 286a												1																																									1	1
977	H'	266 291												1																																									1	1
978	H'	249 304												1																																									1	1
979	H'	134 311												1																																									1	1
980	I	129 223 270 291 311 319																																			1																		1	1
981	IXW	189 223 245 304																																			1																		1	1
982	IXW	214 223																																			1																		1	1
983	C	148 223 288 298 327																																			1																		1	1
984	H'	183c 189 261 278																																			1																		1	1
985	H'	215 311																																			1																		1	1
986	H'	269 325																																			1																		1	1
987	H'	167 243 249																																			1																		1	1
988	H'	216																																			1																		1	1
989	H'	183c 183i 189 218 328a																																			1																		1	1
990	J'	126 147 242 311																																			1																		1	1
991	H'	217 309																																			1																		1	1
992	J'	126 231 266																																			1																		1	1
993	H'	243 245a 311																																			1																		1	1
994	T	126 259g 270 294 296 311																																			1																		1	1
995	IXW	124 160 223 240c																																			1																		1	1
996	other	129 182c 183c 189 207 234																																			1																		1	1
997	C	129 223 298 319 327																																			1																		1	1
998	L1a	148 172 187 188a 189 214 223 230 234 311																																			1																		1	1
999	J1b	126 145 222 261 290																																			1																		1	1
1000	X	183c 189 223 248 278																																			1																		1	1
1001	J1	126 145 261 287																																			1																		1	1
1002	J2	126 193 274																																			1																		1	1
1003	N	129 147a 248																																			1																		1	1
1004	K	224 305 311																																																					1	1	1
1005	K	172 224 246t 311																																																					1	1	1
1006	K	169 224 245 311 335																																																					1	1	1
1007	H'	286 291																																																					1	1	1
1008	I	129 223 297																																																					1	1	1
1009	I	129 145 223 225																																																					1	1	1
1010	U5a1a	256 270 304																																																					1	1	1
1011	U5a	192 249 270 309																																																					1	1	1
1012	U5a1	192 235 256 270 294																																																					1	1	1
1013	H'	129 166t 192 256 266 294																																																					1	1	1
1014	T3	126 145 169 257 292 294																																																					1	1	1
1015	T2	126 218 287 294 296 304																																																					1	1	1
1016	H'	145 180 192 256 320																																																					1	1	1
1017	U5a1	148 192 256 270 291																																																					1	1	1
1018	U5b	129 182c 183c 189 270																																																					1	1	1
1019	H'	182c 183c 189 290																																																					1	1	1
1020	U2	129c 145 183c 189 272																																																					1	1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bi	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pi	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
1021	X	148 183c 189 223 278																																											1							1	1
1022	T1	126 163 171 186 189 294																																													1							1	1
1023	U2	129c 183d 189 256																																													1							1	1
1024	T	126 183d 184d 189 294 296 298																																													1							1	1
1025	H'	183d 184d 189																																													1							1	1
1026	T1	126 163 183 189 294																																													1							1	1
1027	L1c	129 183i 189 223 278 284 294 311																																													1							1	1
1028	X	129 183i 189 223 278																																													1							1	1
1029	X	183i 189 223 255 278																																													1							1	1
1030	H'	176 183i 184i 189																																													1							1	1
1031	T1	126 163 186 189 249 294																																													1							1	1
1032	V	183i 184i 189 239 298																																													1							1	1
1033	W	214 223 292																																													1							1	1
1034	H'	311 322																																													1							1	1
1035	H'	166c																																													1							1	1
1036	V	248 298																																													1							1	1
1037	H'	274 293																																													1							1	1
1038	H'	286 311																																													1							1	1
1039	H'	240c 311																																													1							1	1
1040	U2	129c 183 189																																													1							1	1
1041	H'	301 309																																													1							1	1
1042	H'	169 235 304																																													1							1	1
1043	J'	126 234																																																	1					1	1
1044	I	129 171 223 293c 296i 311																																																	1					1	1
1045	V	231 298																																																	1					1	1
1046	H'	240 261																																																	1					1	1
1047	T1	126 163 186 189 287 294																																																	1					1	1
1048	T	126 219 294																																																	1					1	1
1049	H'	207 225 278																																																	1					1	1
1050	J1b1	126 145 172 192 222 260 261																																																	1					1	1
1051	H'	129 189 256																																																	1					1	1
1052	H'	255																																																	1					1	1
1053	H'	129 180 256																																																				1			1	1
1054	W	223 284 286 292																																																				1			1	1
1055	H'	187 189 207																																																				1			1	1
1056	IXW	129c 189 233																																																				1			1	1
1057	U5a1	192 256 270 311 320																																																				1			1	1
1058	H'	189 209 294 300																																																				1			1	1
1059	H'	189 209 256 294 300																																																				1			1	1
1060	T	126 172 192 294																																																				1			1	1
1061	N1a	147a 223 248 319 320																																																				1			1	1
1062	V	290 298																																																				1			1	1
1063	J'	126 249 288																																																				1			1	1
1064	U1	189 249 288																																																				1			1	1
1065	H'	148 224																																																				1			1	1
1066	H'	130 257																																																				1			1	1
1067	J'	126 188																																																				1			1	1
1068	K	224 288 311													1																																											1	1
1069	H'	209 221													1																																											1	1
1070	T2	126 189 266 274 294 304					1																																																			1	1
1071	T2	126 294 296 304 311					1																																																			1	1
1072	T	126 294 311					1																																																		1	1
1073	H'	162 172					1																																																		1	1
1074	V	162 298					1																																																		1	1
1075	U5b	189 192 217 234 270					1																																																		1	1
1076	H'	234 293					1																																																	1	1
1077	H'	258 311											1																																											1	1
1078	J1b1	126 145 172 231 261											1																																											1	1
1079	IXW	192 223 274 301																																																				1	1	1
1080	J'	126 293c																																																			1	1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
1081	H'	129 192 311																																																		1	1	1
1082	H'	183c 189 239																																																		1	1	1
1083	X	189a 223 271 278																																																		1	1	1
1084	T2	126 192 260 294 304																																																		1	1	1
1085	T2	126 192 207 294 304																																																		1	1	1
1086	T1	126 163 186 189 193d 294																																																		1	1	1
1087	T1	126 163 186 193d 239 294																																																		1	1	1
1088	H'	167 169 274 304																																																1		1	1
1089	H'	129 209 256																																																1		1	1
1090	H'	129 153																																																1		1	1
1091	H'	129 249																																																1		1	1
1092	H'	222 261																																																1		1	1
1093	H'	286																																																1		1	1
1094	J'	126 169 239																																																1		1	1
1095	J'	126 179																																																1		1	1
1096	J'	126 318t																																																1		1	1
1097	K	224 266 291 294 311																																																1		1	1
1098	K	224 278 291 311																																																1		1	1
1099	U2	129c 181 189 260																																																1		1	1
1100	H'	136																							1																											1	1	1
1101	H'	150																																																1				1	1	1
1102	H'	153																	1																																			1	1	1
1103	H'	173		1																																																	1	1	1
1104	H'	217																																						1													1	1	1
1105	H'	220																																				1															1	1	1
1106	H'	289																		1																																		1	1	1
1107	IXW	124 192 223 325																																							1												1	1	1
1108	IXW	124 223 245 278																																							1												1	1	1
1109	IXW	124 223 256																																							1												1	1	1
1110	IXW	124 223 311																																							1												1	1	1
1111	V	124a 153 298																						1																													1	1	1
1112	T1	126 129 163 172 186 189 211 294																																				1															1	1	1
1113	T2	126 129 294 296 304																								1																											1	1	1
1114	other	126 136 223															1																																				1	1	1
1115	T1	126 138 163 186 189 294 311																																																		1				1	1	1
1116	N1b	126 145 176g 223																							1																													1	1	1
1117	J1a	126 145 189 261																								1																													1	1	1
1118	J1b	126 145 222 260 261																1																																					1	1	1
1119	J1b	126 145 222 261 287																																							1													1	1	1
1120	J1	126 145 235 261 291																																								1												1	1	1
1121	J'	126 145 245																	1																																					1	1	1
1122	J1	126 145 261 274 287																									1																											1	1	1
1123	T3	126 146 169 292 294																																								1													1	1	1
1124	T3	126 146 284 292 294 296																	1																																					1	1	1
1125	T2	126 147 224 294 296 297 304																																				1															1	1	1
1126	J'	126 147 242																								1																											1	1	1
1127	T5	126 153 189 257 294																																								1													1	1	1
1128	T3	126 162 186 189 258 292 294 311																			1																																		1	1	1
1129	T1	126 163 179 189 243 294																																	1																			1	1	1
1130	T1 (1)	126 163 185 189																						1																															1	1	1
1131	T1 (1)	126 163 186 189 209																					1																															1	1	1
1132	T1	126 163 186 189 209 278 294																																							1													1	1	1
1133	T1	126 163 186 189 209 294																																						1													1	1	1
1134	T1	126 163 186 189 213 239 294																			1																																	1	1	1
1135	T1	126 163 186 189 248 266 271 294 311																																							1													1	1	1
1136	T1	126 163 186 189 249 294 302																																						1													1	1	1
1137	T1	126 163 186 189 274 294																																							1												1	1	1
1138	T1	126 163 186 189 294 296																								1																											1	1	1
1139	T1 (1)	126 163 186 189 311																						1																													1	1	1
1140	T	126 163 294																																				1															1	1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
1141	T	126 163 294 299																																			1																1	1
1142	J'	126 168																							1																													1	1
1143	J2	126 180 193 278																								1																												1	1
1144	T	126 181 294																									1																											1	1
1145	T	126 183d 189																		1																																		1	1
1146	T	126 184 248 294 296																																			1																1	1
1147	J'	126 186																	1																																			1	1
1148	T1 [1]	126 186 189																		1																																		1	1
1149	J'	126 186 320																	1																																			1	1
1150	J'	126 187 209																																			1																1	1
1151	J1	126 187 261																																			1																1	1
1152	T2	126 189 209 256 294 296																																					1															1	1
1153	X	126 189 223 278																																	1																			1	1
1154	T2	126 189 241 294 296 304																																																						1	1
1155	U1?	126 189 249																								1																													1	1
1156	T	126 192 294 296																																			1																1	1
1157	J2	126 193 304																																			1																1	1
1158	T2	126 207 294 296 304 311																	1																																				1	1
1159	J'	126 214																																			1																1	1
1160	other	126 223 298																				1																																1	1
1161	T3	126 231 292 294 296		1																																																		1	1
1162	T	126 239 294 296																																					1														1	1
1163	J'	126 245																						1																													1	1
1164	T	126 245 294 296																																			1																1	1
1165	T	126 256 290 294 296																																					1														1	1
1166	T	126 256 294 296 324																																						1													1	1
1167	J'	126 260														1																																						1	1
1168	T	126 260 294 296 325																	1																																				1	1
1169	T2	126 272 294 296 304																																			1																1	1
1170	T	126 287 294																							1																													1	1
1171	T	126 288 294 296																								1																													1	1
1172	J'	126 290																																			1																1	1
1173	J'	126 291 319																				1																																	1	1
1174	T3	126 292 294 296 304 311																								1																													1	1
1175	T3	126 292 294 318																																			1																1	1
1176	N1b	128g 145 176g 223 311																																			1																1	1
1177	H'	129 140																	1																																			1	1
1178	N1b	129 145 176g 223																																																				1	1
1179	N1b	129 145 176g 223 249																									1																											1	1
1180	I	129 145 223 311																																			1																1	1
1181	I	129 148 223																							1																													1	1
1182	I	129 172 223 250 311																																																			1	1
1183	I	129 172 223 274 311 319		1																																																	1	1
1184	H'	129 172 311																						1																													1	1
1185	U7	129 181 183d 189 318t																		1																																			1	1
1186	H'	129 183 189 222 311																																			1																1	1
1187	H'	129 183 189 311																																			1																1	1
1188	H'	129 183c 189																				1																																	1	1
1189	IXW	129 183c 189 223 249 311																																					1															1	1
1190	U1	129 183c 189 249 288																																					1															1	1
1191	IXW	129 185 189 223 249 311																																	1																			1	1
1192	I	129 186 223 260																			1																																	1	1
1193	L	129 189 194del 195del 223 224 289 290																							1																													1	1
1194	H'	129 189 234																																			1																1	1
1195	H'	129 189 234 311																								1																												1	1
1196	U2	129 189 249 264																																			1																1	1
1197	H'	129 189 249 311																																																			1	1
1198	U5b	129 189 262a 270																																																			1	1
1199	U5b	129 189 270																																																			1	1
1200	I	129 223 257a 261																																																			1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
1201	I	129 223 264 265 270 311 319																							1																													1	1
1202	I	129 223 294																																											1										1	1
1203	H'	129 248 311																																			1																	1	1
1204	H'	129 287 311																																			1																	1	1
1205	V	129 298 311																																										1										1	1
1206	U2	129c 179 189 311																								1																													1	1
1207	U2	129c 189 214																																										1											1	1
1208	IXW	129c 189 223 287																																											1											1	1
1209	U2	129c 289 256																	1																																					1	1
1210	IXW	133 223															1																																							1	1
1211	H'	134 154 209																1																																							1	1
1212	H'	134 209															1																																							1	1
1213	H'	136 176 218 278																																										1											1	1
1214	H'	136 176 278																																										1											1	1
1215	IXW	136 182c 183c 189 223															1																																							1	1
1216	H'	140 193d 293 311																		1																																					1	1
1217	H'	140 265 311																	1																																					1	1
1218	H'	144 172 189 311																			1																																			1	1
1219	U5b1	144 209 189 270																	1																																					1	1
1220	IXW	145 176 209 223																		1																																					1	1
1221	IXW	145 176 223 258 291																																											1											1	1
1222	N	145 176g 223 258 291																																			1																	1	1
1223	N	145 176g 223 297 311																																			1																	1	1
1224	N	145 176g 223 311																																	1																			1	1
1225	IXW	145 223 256 325del 327																																									1											1	1
1226	N1a	147 172 223 248 278																																			1																	1	1
1227	W	147 223 286 292																																												1									1	1
1228	K	147 224 311	1																																																					1	1
1229	H'	147 234 296																																										1											1	1
1230	U5	147 270																						1																															1	1
1231	N1a	147a 172 223 248																																			1																	1	1
1232	N1a	147a 172 223 248 294 320																	1																																					1	1
1233	N1a	147a 172 223 248 320																	1																																					1	1
1234	H'	147c 172 189 223 270																				1																																	1	1
1235	N1a	147g 172 223 224 248																																	1																			1	1
1236	IXW	148 223 245 270a																																									1											1	1
1237	V	148 261 298		1																																																		1	1
1238	H'	148 311																							1																													1	1
1239	H'	150 189																																																					1	1
1240	H'	150 189 311																								1																													1	1
1241	H'	150 192																																																					1	1
1242	IXW	153 223																						1																															1	1
1243	IXW	153 223 319																																			1																	1	1
1244	V	153c 298															1																																							1	1
1245	H'	162 163																		1																																					1	1
1246	H'	162 213 266																									1																													1	1
1247	H'	162 259																		1																																					1	1
1248	H'	162 274																			1																																					1	1
1249	H'	162 319		1																																																				1	1
1250	T1 [2]	163 186 188 294																				1																																			1	1
1251	T1 [2]	163 189 243 294																		1																																					1	1
1252	H'	164t 311																																			1																	1	1
1253	H'	166 189 278 293 311																	1																																					1	1
1254	H'	166del 311																																																				1	1
1255	U5a1	168 192 256 261 270 304																		1																																					1	1
1256	IXW	169 188 223 239 278 294 309																																										1											1	1
1257	K	169 224 311																																	1																			1	1
1258	H'	169 311																						1																															1	1
1259	H'	171del 304																																										1											1	1
1260	B?	172 187 189 223 320																									1																													1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
1261	H'	172 189																		1																																	1	1
1262	IXW	172 189 223 320																																						1															1	1
1263	U6b	172 219 261 311																								1																													1	1
1264	U6a	172 219 278																																						1															1	1
1265	IXW	172 223																				1																																	1	1
1266	IXW	172 223 248 256 320																				1																																	1	1
1267	K	172 224 231 311																																	1																			1	1
1268	H'	172 248																							1																													1	1
1269	H'	173 260																																	1																			1	1
1270	H'	175 254																																						1															1	1
1271	K	176 224 299 311																		1																																				1	1
1272	H'	176 256 311																																						1															1	1
1273	H'	176del																	1																																					1	1
1274	H'	177 217																																						1															1	1
1275	H'	178 189 234 324																																					1																1	1
1276	H'	178 291 304																																							1														1	1
1277	H'	179 183c 189																																						1															1	1
1278	H'	179 189																																						1															1	1
1279	IXW	179 215 223 256a 284 311																																							1														1	1
1280	IXW	182 189 223 298																				1																																	1	1
1281	U1	182c 183c 189 249																																						1															1	1
1282	H'	182c 189																				1																																	1	1
1283	H'	183c 188 189																		1																																					1	1
1284	H'	183c 189 217																																						1															1	1
1285	H'	183c 189 304 311																				1																																	1	1
1286	H'	183c 189 311															1																																						1	1
1287	H'	184 189 246c																																						1															1	1
1288	IXW	184 201 223 265 311																																							1														1	1
1289	H'	184 209																																					1															1	1
1290	H'	184 234 291 294																																							1														1	1
1291	K	184i 188 189 224 311																								1																													1	1
1292	H'	186 192																				1																																	1	1
1293	H'	186 240																																					1															1	1
1294	K	187 224 298 311																								1																													1	1
1295	K	187 224 311																																						1															1	1
1296	H'	188 189																			1																																			1	1
1297	X	189 192 223 278 292																		1																																					1	1
1298	X	189 192 223 278 292 309																																						1															1	1
1299	U5a1	189 192 234 256 270 311																				1																																	1	1
1300	U5a1	189 192 256 270 311																								1																													1	1
1301	H'	189 194del 195del 293		1																																																				1	1
1302	U5b	189 197 230 270																								1																													1	1
1303	U5b	189 208del 209del 270																	1																																					1	1
1304	U5b	189 209 270 301																								1																													1	1
1305	H'	189 209del 210del																1																																					1	1
1306	H'	189 210del 272del																1																																					1	1
1307	H'	189 211																																					1															1	1
1308	H'	189 217																																					1															1	1
1309	X	189 217 223 278																								1																													1	1
1310	H'	189 217 261 294 299																																																			1	1
1311	H'	189 218 328a																																					1															1	1
1312	IXW	189 223 239 278 298 318																																																				1	1
1313	X	189 223 261 274 278		1																																																		1	1
1314	IXW	189 223 265t																																							1														1	1
1315	X	189 223 274 278																																							1														1	1
1316	X	189 223 278 290 292 294 309																																							1														1	1
1317	X	189 223 278 294 309																																							1														1	1
1318	X	189 223 278 294 309 317t																																							1														1	1
1319	X	189 223 278 301g																																							1														1	1
1320	X	189 223 278 324 325																																	1																			1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bi	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
1321	U1	189 224 249																																			1																1	1
1322	K	189 224 264 311																																			1																1	1
1323	K	189 224 311 325																																									1										1	1
1324	U5b	189 232 270																								1																											1	1
1325	U1	189 239 249 293del																																						1														1	1
1326	U1	189 249 294 318																																	1																			1	1
1327	U1	189 249 327																																							1													1	1
1328	H'	189 265c 291																							1																												1	1
1329	U5b	189 270 278																		1																																		1	1
1330	H'	189 291g 311																				1																																1	1
1331	H'	189 294 296																				1																																1	1
1332	H'	189 300 325																																																			1	1
1333	U7	189 318t																									1																										1	1
1334	X	189a 223 278																																										1										1	1
1335	H'	192 213																																																			1	1
1336	U5a1	192 217 256 270																	1																																				1	1
1337	IXW	192 223 240c																			1																																		1	1
1338	W	192 223 292																		1																																			1	1
1339	W	192 223 292 295																								1																												1	1
1340	U5a1	192 256 270 284 294																																																			1	1
1341	U5a1	192 256 270 291 318t																																					1																1	1
1342	U5a1	192 256 270 318																							1																														1	1
1343	H'	192 261 289																							1																														1	1
1344	H'	192 301																																										1										1	1
1345	H'	192 304 311																		1																																			1	1
1346	H'	193 278																	1																																				1	1
1347	H'	193c																		1																																			1	1
1348	H'	209 218 328a																																									1											1	1
1349	K	209 224 278 293 311																		1																																				1	1
1350	H'	209 229																				1																																1	1
1351	H'	209 265																																																					1	1
1352	H'	209 304 311																																										1										1	1
1353	W	211 223 292 325																																																					1	1
1354	H'	214a 249 290 327																																										1										1	1
1355	H'	218 291																																												1										1	1
1356	H'	220c 243 292																																																					1	1
1357	IXW	223 227 278																																																					1	1
1358	IXW	223 234 274																																																					1	1
1359	IXW	223 256 311																				1																																		1	1
1360	IXW	223 256 327																				1																																		1	1
1361	W	223 264 292																																		1																			1	1
1362	IXW	223 265t																																																					1	1
1363	IXW	223 270 327																				1																																		1	1
1364	IXW	223 278 286 294 309																								1																														1	1
1365	IXW	223 278 294																								1																														1	1
1366	W	223 290 292																																																					1	1
1367	IXW	223 290 318t																																																					1	1
1368	IXW	223 291 294																						1																																1	1
1369	IXW	223 291 319																				1																																		1	1
1370	IXW	223 294 304																							1																															1	1
1371	C	223 298 327																				1																																			1	1
1372	K	224 242 311																							1																															1	1
1373	K	224 246t 265 311																	1																																							1	1
1374	K	224 265 311																																									1												1	1
1375	K	224 290 311																																									1												1	1
1376	K1	224 311 327																																																					1	1
1377	H'	224 319																																										1										1	1
1378	U5	234 270																								1																												1	1
1379	U7	234 309 318t																							1																														1	1
1380	H'	239 292		1																																																		1	1

No.	hg	haplotype	po	cz	au	sz	bv	gs	gc	gn	dn	nw	fn	sw	ic	sa	ka	vf	es	be	nr	ur	kr	ws	bl	rm	si	ni	sd	tc	sp	pt	fr	bq	gk	sy	tk	ar	ge	pl	ad	gy	ro	os	sc	ok	en	cw	ir	wa	ky	wi	anc	Σ	pops
1381	H'	240 286a																		1																																		1	1
1382	H'	240c 242 304																			1																																			1	1
1383	H'	242c 274 288																					1																																		1	1
1384	H'	249 278 327																																	1																					1	1
1385	H'	249 327																							1																																1	1
1386	U5a1a	256 270 292 294																	1																																							1	1
1387	U5a1a	256 270 295																								1																																1	1
1388	U5a1a	256 270 325																																																								1	1
1389	H'	256 294 304 318 320																									1																																1	1
1390	H'	256 295																																	1																					1	1
1391	H'	257 304																									1																																1	1
1392	H'	260 265t 311		1																																																								1	1
1393	H'	261 291																																																									1	1
1394	H'	262 311																	1																																									1	1
1395	V	265 298																								1																																1	1
1396	H'	265c 292																																																									1	1
1397	H'	266 287																									1																																1	1
1398	H'	266 304																																																									1	1
1399	U5	270 304																																																									1	1
1400	H'	271a																				1																																						1	1
1401	V	274 298																									1																																1	1
1402	H'	278 294																		1																																								1	1
1403	H'	290 295																																																									1	1
1404	U7	290 304 318t																																																								1	1
1405	H'	291 304																																																								1	1
1406	H'	293t 311																																																								1	1
1407	T2 (1)	294 296 299 304																		1																																								1	1
1408	V	298 301 311																							1																																		1	1
1409	H'	300 325																																																								1	1
1410	H'	304g															1																																										1	1
1411	U7	309 318t 319																									1																																1	1
1412	H'	311 316																																																									1	1
	individuals		37	83	116	76	49	200	117	107	48	565	176	60	467	176	83	32	149	55	40	53	18	38	141	48	90	68	69	49	181	54	64	105	125	49	72	191	124	117	50	130	92	106	891	78	258	86	101	92	49	181	48	6454	51
	haplotypes (ht)		21	48	66	39	33	105	69	67	28	207	66	32	102	30	39	22	70	54	37	49	14	28	70	42	43	50	40	37	92	33	40	46	55	42	58	113	75	91	27	36	47	56	219	38	128	44	48	44	22	69	36	1412	51
	unique		15	33	48	23	25	83	54	48	21	141	32	23	52	20	25	16	46	53	34	46	11	21	53	38	31	41	31	33	69	27	30	31	36	39	48	74	61	79	20	24	31	32	125	21	101	33	35	31	13	37	27	939	51
	multiple hts		6	15	18	16	8	22	15	19	7	66	34	9	50	10	14	6	24	1	3	3	3	7	17	4	12	9	9	4	23	6	10	15	19	3	10	39	14	12	7	12	16	24	94	16	27	11	13	13	9	32	9	473	51
	shared hts		20	36	43	31	26	69	45	37	21	117	46	24	61	17	31	16	49	33	27	28	12	18	44	21	28	17	23	23	64	23	27	33	40	26	34	60	41	46	22	20	35	35	132	31	84	29	35	31	18	51	26		
	private hts		1	12	23	8	7	36	24	30	7	90	20	8	41	13	8	6	21	21	10	21	2	10	26	21	15	33	17	14	28	10	13	13	15	16	24	53	34	45	5	11	12	21	87	7	44	15	13	13	4	18	10		
	% unique hts		71	69	73	59	76	79	78	72	75	68	48	72	51	67	64	73	66	98	92	94	79	75	76	90	72	82	78	89	75	82	75	67	65	93	83	65	81	87	74	67	66	57	57	55	79	75	73	70	59	54	75		
	% multiple hts		29	31	27	41	24	21	22	28	25	32	52	28	49	33	36	27	34	2	8	6	21	25	24	10	28	18	23	11	25	18	25	33	35	7	17	35	19	13	26	33	34	43	43	45	21	25	27	30	41	46	25		
	% shared hts		95	75	65	79	79	66	65	55	75	57	70	75	60	57	79	73	70	61	73	57	86	64	63	50	65	34	58	62	70	70	68	72	73	62	59	53	55	51	81	65	74	63	60	82	66	66	73	70	82	74	72		
	% private hts		5	25	35	21	21	34	35	45	25	43	30	25	40	43	21	27	30	39	27	43	14	36	37	50	35	66	43	38	30	30	33	28	27	38	41	47	45	49	19	35	26	38	40	18	34	34	27	30	18	26	28		

CODES

unique hts — occur only once within a population,

multiple hts — occur more than once within a popultion

shared hts — shared with other populations

private hts — occur *exclusively* in one population (can be either unique or multiple, this latter are indicated in **bold type**)

(1) probable reversion at 294

(2) probable reversion at 126

(3) probable reversion at 270

References

Abbaszadegan M, Huber MS, Gerba CP & Pepper IL (1993) Detection of Enteroviruses in groundwater with the Polymerase Chain Reaction. *Applied and Environmental Microbiology* **59**, 1318-1324.

Adachi J & Hasegawa M (1995) Improved dating of the Human-Chimpanzee separation in the mitochondrial DNA tree – heterogeneity among amino-acid sites. *Journal of Molecular Evolution* **40**, 622-628.

Adcock GJ, Dennis ES, Easteal S, Huttley GA, Jermiin LS, Peacock WJ & Thorne A (2001) Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins. *Proceedings of the National Academy of Sciences USA* **98**, 537-542.

Albert AM (1998) The use of vertebral ring epiphyseal union for age estimation in two cases of unknown identity. *Forensic Science International* **97**, 11-20.

Allen M, A. E, Meyers S, Handt O, Saldeen T, von Haeseler A, Pääbo S & Gyllensten U (1998) Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. *Journal of Forensic Science* **43**, 453-464.

Ammerman AJ (1989) On the Neolithic transition in Europe: a comment on Zvelebil & Zvelebil (1988). *Antiquity* **LXIII**, 162-165.

Ammerman AJ & Cavalli-Sforza LL (1971) Measuring the rate of spread of early farming in Europe. *Man* **6**, 674-688.

Amos W & Harwood J (1998) Factors affecting levels of genetic diversity in natural populations. *Philosophical Transactions of the Royal Society of London. Series B.* **353**, 177-186.

Anderson S, Bankier A, Barrel B, Debruijn M, Coulson A, Drouin J, Eperon I, Nierlich D, Roe B, Sanger F, Schreier P, Smith A & Young I (1981) Sequence organisation of the human mitochondrial genome. *Nature* **290**, 457-465.

Ankel-Simons F & Cummins JM (1996) Misconceptions about mitochondria and mammalian fertilization: Implications for theories on human evolution. *Proceedings of the National Academy of Science USA* **93**, 13859-13863.

Anslinger K, Weichhold G, Keil W, Bayer B & Eisenmenger W (2001) Identification of the skeletal remains of Martin Bormann by mtDNA analysis. *International Journal of Legal Medicine* **114**, 194-196.

Aquadro CF & Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics* **103**, 287-312.

Aris-Brosou S & Excoffier L (1996) The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Molecular Biology and Evolution* **13**, 494-504.

Arnold J (1984) *From Roman Britain to Saxon England.* London.

Avise J, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel J, Reeb C & Saunders N (1987) Intraespecific phylogeography: the molecular bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Awadalla P, Eyre-Walker A & Smith JM (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* **286**, 2524-2525.

Awadalla P, Eyre-Walker A & Smith JM (2000) Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**, 1931a.

Baasner A, Schafer C, Junge A & Madea B (1998) Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. *Forensic Science International* **98**, 169-178.

Bachmann L, Scholz M, Broghammer M, Giddings I & Pusch CM (2000) Voltage-induced release of nucleic acids from palaeontological samples. *Electrophoresis* **21**, 1488-1492.

Bada JL, Wang XS, Poinar HN, Pääbo S & Poinar GO (1994) Amino-acid racemization in amber-entombed insects - Implications for DNA preservation. *Geochimica et Cosmochimica Acta* **58**, 3131-3135.

Bailey JF, Richards M, Macaulay V, Colson I, James IT, Bradley DG, Hedges RE & Sykes B (1996) Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. *Proceedings of the Royal Society of London B* **22**, 1467-1473.

Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BVR, Rasanayagam A & Hammer MF (1998) Female gene flow stratifies Hindu castes. *Nature* **395**, 651-652.

Bamshad MJ, Watkins WS, Rogers AR, Harpending HC, Soodyall H, Kere J, Fraley AE, Krakowiak P, Sung S & Jorde LB (1997) DNA variation and the evolution of modern humans. *American Journal of Human Biology* **9**, 103-103.

Bandelt HJ, Forster P & Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Bandelt HJ, Forster P, Sykes BC & Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743-753.

Bandelt HJ, Lahermo P, Richards M & Macaulay V (2001) Detecting errors in mtDNA data by phylogenetic analysis. *International Journal of Legal Medicine* **115**, 64-69.

Barbujani G & Bertorelle G (2001) Genetics and the population history of Europe. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 22-25.

Barbujani G, Bertorelle G & Chikhi L (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. *American Journal of Human Genetics* **62**, 488-491.

Barbujani G & Sokal RR (1990) Zones of sharp genetic change in Europe are also Linguistic boundaries. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 1816-1819.

Barbujani G, Sokal RR & Oden NL (1995) Indo-European origins - a computer-simulation test of 5 hypotheses. *American Journal of Physical Anthropology* **96**, 109-132.

Barker RE (1985) *Prehistoric farming in Europe*. Cambridge.

Baron H, Hummel S & Herrmann B (1996) *Mycobacterium tuberculosis* Complex DNA in ancient human bones. *Journal of Archaeological Sciences* **23**, 1-5.

Bede (731) *Ecclesiastical History of the English People*.

Belyaeva OV, Mikulich AI, Slominsky PA & Limborska SA (2000) Mitochondrial DNA variations in eastern Slavonic populations. NCBI (direct submission), AY005336-AY005390.

Bendall KE & Sykes BC (1995) Length Heteroplasmy in the first hypervariable segment of the Human mtDNA control region. *American Journal of Human Genetics* **57**, 248-256.

Bender K, Schneider PM & Rittner C (2000) Application of mtDNA sequence analysis in forensic casework for the identification of human remains. *Forensic Science International* **113**, 103-107.

Béraud-Colomb E, Roubin R, Martin J, Maroc N, Gardeisen A, Trabuchet G & Goosens M (1995) Human β-Globin gene polymorphisms characterized in DNA extracted from ancient bones 12,000 years old. *American Journal of Human Genetics* **57**, 7-15.

Bermisheva MA, Belyaeva OV, Slominsky PA, Khusnutdinova EK & Limborska SA (2000) Mitochondrial DNA diversity in eastern Slavonic populations. NCBI (direct submission), AF292943-AF292995.

Bertorelle G & Barbujani G (1995) Analysis of DNA diversity by spatial autocorrelation. *Genetics* **140**, 811-819.

Bertranpetit J & Cavalli-Sforza LL (1991) A genetic reconstruction of the history of the population of the Iberian Peninsula. *Annals of Human Genetics* **55**, 51-67.

Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P & Comas D (1995) Human mitochondrial DNA variation and the origin of Basques. *Annals of Human Genetics* **59**, 63-81.

Bjerck HB (1995) The North Sea Continent and the pioneer settlement of Norway. In *Man & sea in the Mesolithic - Coastal settlement above and below present sea level. Proceedings of the International Symposium, Kalundborg, Denmark 1993*, pp. 129-134 [A Fischer, editor]. Oxford: Oxbow Monographs.

Bodmer WF (1993) The genetics of the Celtic population. *Proceedings of the British Academy* **82**, 37-57.

Bolton J (1996) The world upside down. Plague as an agent of economic and social change. In *The Black Death in England*, pp. 17-78 [WM Ormrod and PG Lindley, editors]. Stamford: Paul Watkins.

Boom R, Sol CJA, Salimans MM, Jansen CL, Wertheim-van Dillen PM & van der Noorda J (1990) Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology* **28**, 495-503.

Bradley R (1978) *The prehistoric settlement of Europe*. London: Routledge and Kegan Paul.

Breimer LH (1990) Molecular mechanisms of oxygen radical carcinogenesis and mutagenesis - the role of DNA-base damage. *Molecular Carcinogenesis* **3**, 188-197.

Breimer LH & Lindahl T (1985) Enzymatic Excision of DNA Bases Damaged by Exposure to Ionizing-Radiation or Oxidizing-Agents. *Mutation Research* **150**, 85-89.

Brkic H, Strinovic D, Kubat M & Petrovecki V (2000) Odontological identification of human remains from mass graves in Croatia. *International Journal of Legal Medicine* **114**, 19-22.

Bromham L, Eyre-Walker A, Smith NH & Smith JM (2003) Mitochondrial Steve: paternal inheritance of mitochondria in humans. *Trends in Ecology & Evolution* **18**, 2-4.

Brothwell DR (1981) *Digging up bones: the excavation, treatment and study of human skeletal remains*, 3rd. ed. London: British Museum.

Brown BW & Russell K (1996) Multiplicity program: Identifying important results from multiple statistical tests - Version 2.0: University of Texas, Houston.

Brown MD, Hosseini SH, Torroni A, Bandelt H-J, Allen JC, Schurr TG, Scozzari R, Cruciani F & Wallace DC (1998) mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? *American Journal of Human Genetics* **63**, 1852-1861.

Brown TA, Allaby RG, Brown KA, O'Donoghue K & Sallares R (1994) DNA in wheat seeds from European archaeological sites. *Experientia* **50**, 571-575.

Brown WM (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proceedings of the National Academy of Science USA* **77**, 3605-3609.

Brzustowski J (2001) The rarefaction calculator. Available 21 June 2001 at http://www2.biology.ualberta.ca.jbrzusto/rarefct.php.

Budd P, Chenery C, Montgomery J, Evans J & Powlesland D (2002) Anglo-saxon residential mobility at west Heslerton, North Yorkshire, UK from combined O- and Sr-isotope analyses. *8th International Conference on Plasma Source Mass Spectrometry.*

Budowle B, Allard MW & Wilson MR (2002a) Critique of interpretation of high levels of heteroplasmy in the human mitochondrial DNA hypervariable region I from hair. *Forensic Science International* **126**, 30-33.

Budowle B, DiZinno JA & Wilson M (2002b) Interpretation guidelines for mitochondrial DNA sequencing. Washington DC: Federal Bureau of Investigation.

Butler JM & Reeder DJ (2003) Short Tandem Repeat DNA internet database. Available at http://www.cstl.nist.gov/biotech/strbase/.

Calafell F, Underhill P, Tolun A, Angelicheva D & Kalaydjieva L (1996) From Asia to Europe: Mitochondrial DNA sequence variability in Bulgarians and Turks. *Annals of Human Genetics* **60**, 35-49.

Campbell JB (1986) Hiatus and continuity in the British Upper Palaeolithic: a view from the Antipodes. In *Studies in the Upper Palaeolithic of Britain and Northwest Europe*, pp. 7-42 [DA Roe, editor]. Oxford: BAR International Series.

Campbell M (1984) Population pressure inheritance and the land market in a 14th century peasant community. In *Land, kinship and life cycle.*, pp. 87-134 [RM Smith, editor]. Cambridge: Cambridge University Press.

Cann RL, Brown WM & Wilson A (1984) Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* **106**, 479-499.

Cann RL, Stoneking M & Wilson AS (1987) Mitochondrial DNA and human evolution. *Nature* **325**, 31-36.

Cano RJ, Poinar HN, Pieniazek NJ, Acra A & Poinar GO (1993) Amplification and sequencing of DNA from a 120-135-million-year-old weevil. *Nature* **363**, 536-538.

Capelli C, Tschentscher F & Pascali VL (2003) "Ancient" protocols for the crime scene?. Similarities and differences between forensic genetics and ancient DNA analysis. *Forensic Science International* **131**, 59-64.

Carbonell E, Bermúdez de Castro JM, Arsuaga JL, Diez JC, Rosas A, Cuenca-Bescos G, Sala R, Mosquera M & Rodríguez P (1995) Lower Pleistocene hominids and artifacts from Atapuerca-TD6 (Spain). *Science* **269**, 826-827.

Case H (1969) Neolithic explanations. *Antiquity* **XLIII**, 176-186.

Cattaneo C, Smillie DM, Gelsthorpe K, Piccinini A, Gelsthorpe AR & Sokol RJ (1995) A simple method for extracting DNA from old skeletal material. *Forensic Science International* **74**, 167-174.

Cavalli-Sforza F & Edwards AWF (1967) Phylogenetic analyses: models and estimation procedures. *American Journal of Human Genetics* **19**, 233-257.

Cavalli-Sforza LL & Cavalli-Sforza F (1995) *The great human diasporas: the history of diversity and evolution*: Addison-Wesley Publishing Company.

Cavalli-Sforza LL, Menozzi P & Piazza A (1993) Demic expansions and human evolution. *Science* **259**, 639-346.

Cavalli-Sforza LL, Menozzi P & Piazza A (1994) *History and Geography of Human Genes*. Princeton.

Cavalli-Sforza LL & Minch E (1997) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *American Journal of Human Genetics* **61**, 247-251.

Cavelier L, Jazin E, Jalonen P & Gyllensten U (2000) MtDNA substitution rate and segregation of heteroplasmy in coding and noncoding regions. *Human Genetics* **107**, 45-50.

Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AC & Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *American Journal of Human Genetics* **57**, 133-149.

Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V & Barbujani G (1998a) Clines of nuclear DNA markers suggest a largely neolithic ancestry of the European gene pool. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 9053-9058.

Chikhi L, Destro-Bisol G, Pascali V, Baravelli V, Dobosz M & Barbujani G (1998b) Clinal variation in the nuclear DNA of Europeans. *Human Biology* **70**, 643-657.

Cipollaro M, Di Bernardo G, Galano G, Galderisi U, Guarino F, Angelini F & Cascino A (1998) Ancient DNA in human bone remains from Pompeii archaeological site. *Biochemical and Biophysical Research Communications* **247**, 901-904.

Clark AG & Lindly JM (1991) On paradigmatic biases and Paleolithic research traditions. *Current Anthropology* **32**, 577-587.

Clark JGD (1932) *The Mesolithic Age in Britain*. Cambridge: Cambridge University Press.

Clark JGD (1936) *The Mesolithic settlement of northern Europe: a study of the food-gathering peoples of northern Europe during the early post-glacial period*. Cambridge: Cambridge University Press.

Clark JGD (1965a) Radiocarbon dating and the expansion of farming culture from the Near East over Europe. *Proceedings of the Prehistoric Society* **31**, 58-73.

Clark JGD (1965b) Radiocarbon dating and the spread of farming economy. *Antiquity* **XXXIX**, 45-48.

Clark JGD (1975) *The earlier Stone Age settlement of Scandinavia*. Cambridge: Cambridge University Press.

Clifford J (1989) *Eyam Plague, 1665-1666*. Sheffield.

Cohen Salama M & Equipo Argentino de Antropología Forense (1992) *Tumbas anonimas - Informe sobre la identificación de restos de víctimas de la represión ilegal*. Buenos Aires.

Coles BJ (1998) Doggerland: a speculative survey. *Proceedings of the Prehistoric Society* **64**, 45-81.

Coles BJ (1999) Doggerland's loss and the Neolithic. In *Bog bodies, sacred sites and wetland archaeology*, pp. 51-57 [BJ Coles, J Coles and M Schou Jørgensen, editors]. Exeter: WARP ocassional paper.

Colson IB, Bailey JF, Vercauteren M, Sykes B & Hedges RE (1997) The preservation of ancient DNA and bone diagenesis. *Ancient Biomolecules* **1**, 109-117.

Comas D, Calafell F, Mateu E, Pérez-Lezaun A & Bertranpetit J (1996) Geographic variation in human mitochondrial DNA control region sequence: The population history of Turkey and its relationship to the European populations. *Molecular Biology and Evolution* **13**, 1067-1077.

Comas D, Calafell F, Mateu E, Pérez-Lezaun A & Bertranpetit J (1998a) HLA evidence for the lack of genetic heterogeneity in Basques. *Annals of Human Genetics* **62**, 123-132.

Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E & Bertranpetit J (1997) Mitochondrial DNA variation and the origin of the Europeans. *Human Genetics* **99**, 443-449.

Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Martinez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D, Pettener D & Bertranpetit J (1998b) Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *American Journal of Human Genetics* **63**, 1824-1838.

Comps B, Gomory D, Letouzey J, Thiebaut B & Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics* **157**, 389-397.

CONADEP (1984) *Nunca más - Informe de la Comisión Nacional sobre la desaparición de personas*. Buenos Aires: EUDEBA.

Cooper A (1994) Ancient DNA sequences reveal unsuspected phylogenetic relationships within New Zealand Wrens (*Acanthisittidae*). *Experientia* **50**, 558-563.

Cooper A (1997) Ancient DNA - How do you really know when you have it? Reply. *American Journal of Human Genetics* **60**, 1001-1002.

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J & Ward R (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704-707.

Cooper A & Poinar HN (2000) Ancient DNA: Do it right or not at all. *Science* **289**, 1139-1139.

Cooper L (1996) A Roman cemetery in Newarke Street, Leicester. *Transactions of Leicestershire Archaeological and Historical Society* **LXX**, 1-90.

Corach D, Sala A, Penacino G, Ianucci N, Berbardi P, Doretti M, Fonderbrider L, Ginarte A, Inchaurregui A, Somigliana S, Turner S & Hagelberg E (1997) Additional appraoches to DNA typing of skeletal remains: The search for "missing" persons killed during the last dictatorship in Argentina. *Electrophoresis* **18**, 1608-1612.

CorteReal H, Macaulay VA, Richards MB, Hariti G, Issad MS, CambonThomsen A, Papiha S, Bertranpetit J & Sykes BC (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Annals of Human Genetics* **60**, 331-350.

Creighton C (1894) *A History of epidemics in Britain*. Cambridge: Cambridge University Press.

Cummins JM (2002) The role of maternal mitochondria during oogenesis, fertilization and embryogenesis. *Reproductive BioMedicine Online* **4**, 176-182.

Dansgaard W, Johnson SJ, Clausten HB, Dahl-Jensen D, Gundestrup NS, Hammer CU, Hviber CS, Steffensen JP, Sveinbjörnsdottir AE, Jouzel J & Bond G (1993) Evidence for general instability of past climate from a 250 kyr ice core. *Nature* **364**, 218-219.

Deakin W, Rowley-Conwy P & Shaw CH (1998) Amplification and sequencing of DNA from preserved sorghum (*Sorghum bicolor*) of up to 2,800 years antiquity found at Qasr Ibrim. *Ancient Biomolecules* **2**, 27-41.

Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, Goedert JJ, Buchbinder SP, Vittinghoff E, Gomperts E, Donfield S, Vlahov D, Kaslow R, Saah A, Rinaldo C, Detels R & O'Brien SJ (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. *Science* **273**, 1856-1862.

DeBoer SH, Ward LJ, Li X & Chittaranjan S (1995) Attenuation of PCR inhibition in the presence of plant compounds by addition of BLOTTO. *Nucleic Acids Research* **23**, 2567-2568.

DeGusta D, Cook C & Sensabaugh G (1994) Dentina as source of ancient DNA. *Ancient DNA Newsletter* **2**, 13.

DeGusta M & White TD (1996) On the use of skeletal collections for DNA analyses. *Ancient Biomolecules* **1**, 89-92.

Delghandi M, Utsi E & Krauss S (1998) Saami mitochondrial DNA reveals deep maternal lineage clusters. *Human Heredity* **48**, 108-114.

Denaro M, Blanc H, Johnson MJ, Chen KH, Wilmsen E, Cavalli-Sforza LL & Wallace DC (1981) Ethnic variation in Hpa-1 endonuclease cleavage patterns of human mitochondrial DNA. *Proceedings of the National Academy of Science USA* **78**, 5768-5772.

Dennell R (1983) *European economic prehistory: a new approach.* London.

Derenko MV, Grzybowski T, Malyarchuk BA, Czarny J, Miscicka-Sliwka D & Zakharov IA (2001) The presence of mitochondrial haplogroup X in Altaians from South Siberia. *American Journal of Human Genetics* **69**, 237-241.

Devoto F (2003) *Historia de la Inmigración en la Argentina (History of the immigration in Argentina).* Buenos Aires: Editorial Sudamericana.

Diamond JM (1997) The language steamrollers. *Nature* **389**, 544-546.

Dianov G, Price A & Lindahl T (1992) Generation of single nucleotide repair patches following excision of uracil residues from DNA. *Molecular and Cellular Biology* **12**, 1605-1612.

Dolukhanov PM (2000) Prehistoric revolutions and languages in Europe. In *The roots of peoples and languages of northern Eurasia II and III*, pp. 71-78 [A Künnap, editor]. Tartu: University of Tartu.

Donnelly P & Tavaré S (1986) The ages of alleles and a coalescent. *Advances in Applied probability* **18**, 1-19.

Drancourt M, Aboudharam G, Signoli M, Dutour O & Raoult D (1998) Detection of 400-year-old *Yersinia pestis* in human dental pulp: An approach to the diagnosis of ancient septicemia. *Proceedings of the National Academy of Sciences USA* **95**, 12637-12640.

Drancourt M & Raoult D (2002) Molecular insights into the history of plague. *Microbes and Infection* **4**, 105-109.

Dudar JC (1997) A new strategy for DNA sex determination from ancient human skeletons. *Proceedings of the American Association of Physical Anthropology Conference.*

Dumolin-Lapegue S, Pemonge MH, Gielly L, Taberlet P & Petit RJ (1999) Amplification of oak DNA from ancient and modern wood. *Molecular Ecology* **8**, 2137-2140.

Dyer J (1990) *Ancient Britain*. London: Batsford.

Egger J & Wilson J (1983) Mitochondrial inheritance in a mitochondrially mediated disease. *New England Journal of Medicine* **309**, 142-146.

ElMousadik A & Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree *Argania spinosa* (L) Skeels endemic to Morocco. *Theoretical and Applied Genetics* **92**, 832-839.

Elson JL, Andrews RM, Chinnery PF, Lightowlers RN, Turnbull DM & Howell N (2001) Analysis of European mtDNAs for recombination. *American Journal of Human Genetics* **68**, 145-153.

Ember CR (1978) Myths about hunter-gatherers. *Ethnology* **17**, 439-448.

Erlich HA, Gelfand D & Sninsky JJ (1991) Recent advances in the polymerase chain reaction. *Science* **252**, 1643-1651.

Evans B (1975) *The environment of early man in the British Isles*. Berkeley: University of California Press.

Evison MP (1999) Perspectives on the Holocene in Britain: Human DNA. *Quaternary Proceedings* **7**, 615-623.

Evison MP, Smillie DM & Chamberlain AT (1997) Extraction of single-copy nuclear DNA from forensic specimens with a variety of postmortem histories. *Journal of Forensic Sciences* **42**, 1032-1038.

Evison VI (1987) *Dover: the Buckland Anglo-Saxon cemetery*. London: Historic Buildings and Monuments Commission for England.

Ewen WJ (1972) The sampling theory of selective neutral alleles. *Theoretical Population Biology* **3**, 87-112.

Excoffier L, Smouse PE & Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Eyre-Walker A, Smith NH & Smith JM (1999a) How clonal are human mitochondria? *Proceedings of the Royal Society of London Series B-Biological Sciences* **266**, 477-483.

Eyre-Walker A, Smith NH & Smith JM (1999b) Reply to Macaulay *et al.* (1999): mitochondrial DNA recombination - reasons to panic. *Proceedings of the Royal Society of London Series B-Biological Sciences* **266**, 2041-2042.

Faerman M, Filon D, Kahila G, Greenblatt CL, Smith P & Oppenheim A (1995) Sex identification of archaeological human remains based on amplification of the X and Y amelogenin alleles. *Gene* **167**, 327-332.

Fairbanks RG (1989) A 17,000-year glacio-eustatic sea level record: influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* **342**, 637-642.

Falsetti AB & Sokal RR (1993) Genetic structure of human populations in the British Isles. *Annals of Human Biology* **20**, 215-229.

Faure G (1986) *Principles of isotope geology*. New York: John Wiley & Sons Inc.

Fay JC & Wu C (1999) A human bottleneck can account for discordance between patterns of mitochondrial versus nuclear DNA variation. *Molecular Biology and Evolution* **16**, 1003-1005.

Ferreira LF, Britto C, Cardoso MA, Fernandes O, Reinhard K & Araujo A (2000) Paleoparasitology of Chagas disease revaled by infected tissues from Chilean mummies. *Acta Tropica* **75**, 79-84.

Filon D, Faerman M, Smith P & Oppenheim A (1995) Sequence analysis reveals a β-thalassaemia mutation in the DNA of skeletal remains from the archaeological site of Akhziv, Israel. *Nature Genetics* **9**, 365-368.

Fischer A (1991) Pioneers in deglaciated landscapes: the expansion and adaptation of Late Paleolithic societies in Southern Scandinavia. In *The Late Glacial in north-west Europe: Human adaptation and environmental change at the end of the Pleistocene*, pp. 100-122 [N Barton, AJ Roberts and DA Roe, editors]. Oxford: CBA Research Reports.

Fleischer RC, Olson SL, James HF & Cooper AC (2000) Identification of the extinct Hawaiian eagle (*Haliaeetus*) by mtDNA sequence analysis. *Auk* **117**, 1051-1056.

Forster P, Harding R, Torroni A & Bandelt HJ (1996) Origin and evolution of native American mtDNA variation: A reappraisal. *American Journal of Human Genetics* **59**, 935-945.

Forster P (2003) To err is human. *Annals of Human Genetics* **67**, 2-4.

Fortune J (2002) Stone Age Columbus. Available at http://www.bbc.co.uk/science/horizon/2002/columbustrans.shtml.

Francalacci G (1995) DNA recovery from ancient tissues: problems and perspectives. *Human Evolution* **10**, 81-91.

Francalacci P, Bertranpetit J, Calafell F & Underhill PA (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *American Journal of Physical Anthropology* **100**, 443-460.

Fraser AM (1992) *The gypsies*. Oxford: Blackwell.

Frenzel B, Pécsi M & Velichko AA (1992) *Atlas of Paleoclimates and Paleoenvironments of the northern hemisphere. Late Pleistocene-Holocene*. Budapest: Gustav Fisher Verlag.

Frick H, Leonhardt M & Starck D (1991) *Human Anatomy*. New York: Georg Thieme Verlag.

Gamkrelidze TV & Ivanov VV (1990) The Early History of Indo-European Languages. *Scientific American* **March**, 110-116.

Gerstenberger J (2002) Analyse alter DNA zur Ermittlung von Heiratsmustern in einer frühmittelalterlichten Bevölkerung (Analysis of ancient DNA for the determination of wedding patterns in an early medieval population), Georg-August-Universität Göttingen.

Gilbert MT, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, Lynnerup N & Cooper A (2003a) Characterization of genetic miscoding lesions caused by postmortem damage. *American Journal of Human Genetics* **72**, 48-61.

Gilbert MT, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N & Cooper A (2003b) Distribution patterns of postmortem damage in human mitochondrial DNA. *American Journal of Human Genetics* **72**, 32-47.

Gildas (ca.550) De Excidio Britanniae (On the Ruin of Britain).

Giles RE, Blanc H, Cann HM & Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences USA* **77**, 6715-6718.

Gill P, Ivanov PL, Kimptom C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E & Sullivan K (1994) Identification of the remains of the Romanov family by DNA analysis. *Nature Genetics* **6**, 130-135.

Gimbutas M (1970) Proto-Indo-European culture: The Kurgan culture during the fifth, fourth and third millennia BC. In *Indo-European and Indo-Europeans*, pp. 155-195 [G Cardona, HM Hoenigswald and A Senn, editors]. Philedelphia: University of Pennsylvania State.

Ginther C, Isseltarver L & King M-C (1992) Identifying individuals by sequencing mitochondrial DNA from teeth. *Nature Genetics* **2**, 135-138.

Goddall A, Huddle J & Blackburn M (in prep.) Chapter 4.III Finds. Copper alloy brooches. Ansate and disc brooches. In *Excavation at Norwich Castle 1987-98*. Norfolk: NAU.

Goldberg J (1988) Mortality and economic change in the Diocese of York, 1390-1514. *Northern History* **24**, 38-55.

Goldberg J (1996) Introduction. In *The Black Death in England*, pp. 1-16 [WM Ormrod and PG Lindley, editors]. Stamford: Paul Watkins.

Golenberg EM (1990) Chloroplast DNA sequence from a Miocene *Magnolia* species. *Nature* **344**, 656-658.

Goloubinoff P, Pääbo S & Wilson AC (1993) Evolution of maize inferred from sequence diversity of an Adh2 gene segment from archaeological specimens. *Proceedings of the National Academy of Science of the United Statesof America* **90**, 1997-2001.

González-Oliver A, Márquez-Morfin L, Jiménez JC & Torre-Blanco A (2001) Founding Amerindian mitochondrial DNA lineages in ancient Maya from Xcaret, Quintana Roo. *American Journal of Physical Anthropology* **116**, 230-235.

Greenspoon SA, Scarpetta MA, Drayton ML & Turek SA (1998) QIAamp spin columns as a method of DNA isolation for forensic casework. *Journal of Forensic Sciences* **43**, 1024-1030.

Gresham D, Morar B, Underhill P, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen PD, Tournev I, de Pablo R, Kucinskas V, Pérez-Lezaun A, Marushiakova E, Popov V & Kalaydjieva L (2001) Origins and divergence of the Roma (Gypsies). *American Journal of Human Genetics* **69**, 1314-1331.

Gresham D, Passarino G, Tournev I, de Pablo R, Kucinskas V, Calafell F & Kalaydjieva L (2000) Mitochondrial lineages in the Roma. *American Journal of Human Genetics* **67**, 1275.

Gurven M (2000) How can we distinguish between mutational "hot spots" and "old sites" in human mtDNA samples? *Human Biology* **72**, 455-471.

Gutin JC (1995) Remains in Spain now reign as oldest Europeans. *Science* **269**, 754-755.

Gyllensten U, Wharton D, Josefsson A & Wilson AC (1991) Paternal inheritance of mitochondrial DNA in mice. *Nature* **352**, 255-257.

Haas CJ, Zink A, Palfi G, Szeimies U & Nerlich AG (2000) Detection of leprosy in ancient human skeletal remains by molecular identification of Mycobacterium leprae. *American Journal of Clinical Pathology* **114**, 428-436.

Hagelberg E (2003) Recombination or mutation rate heterogeneity? Implications for mitochondrial Eve. *Trends in Genetics* **19**, 84-90.

Hagelberg E & Clegg JB (1991) Isolation and characterization of DNA from archaeological bone. *Proceedings of the Royal Society of London B* **244**, 45-50.

Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhovel W, Clegg JB & Bowden DK (1999a) Evidence for mitochondrial DNA recombination in a human population of Island Melanesia. *Proceedings of the Royal Society of London B* **266**, 485-492.

Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhovel W, Clegg JB & Bowden DK (2000) Evidence for mitochondrial DNA recombination in a human population of Island Melanesia: correction. *Proceedings of the Royal Society of London B* **267**, 1595-1596.

Hagelberg E, Kayser M, Nagy M, Roewer L, Zimdahl H, Krawczak M, Lio P & Schiefenhovel W (1999b) Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y-chromosome and HLA markers. *Philosophical Transactions of the Royal Society of London. Series B.* **354**, 141-152.

Hagelberg E, Sykes B & Hedges RE (1989) Ancient bone DNA amplified. *Nature* **342**, 485.

Hall RA (1990) *Viking age archaeology in Britain and Ireland.* Shire: Princes Risborough.

Hamerow H (1997) Migration theory and the Anglo-Saxon "identity crisis". In *Migrations and invasions in archaeological explanation*, pp. 33-44 [J Chapman and H Hamerow, editors]. Oxford: BAR International Series.

Handt O (1996) The retrieval of ancient human DNA sequences. *American Journal of Human Genetics* **59**, 369-376.

Handt O, Hoss M, Krings M & Pääbo S (1994a) Ancient DNA: methodological challenges. *Experimentia* **50**, 524-529.

Handt O, Meyer S & von Haesler A (1998) Compilation of human mtDNA control region sequences. *Nucleic Acids Research* **26**, 126-129.

Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, Georgiev O, Bauer K, Stone A, Hedges RE, Schaffner W, Utermann G, Sykes B & Pääbo S (1994b) Molecular genetic analyses of the Tyrolean Ice man. *Science* **264**, 1775-1778.

Hansen AJ, Willerslev E, Wiuf C, Mourier T & Arctander P (2001) Statistical evidence for miscoding lesions in ancient DNA templates. *Molecular Biology and Evolution* **18**, 262-265.

Härke H (1990) Warrior Graves - the Background of the Anglo-Saxon Weapon Burial Rite. *Past & Present* **126**, 22-43.

Härke H (1995) Chapter 4: Discussion of the finds. Weapon burials and knives. In *Two Oxfordshire Anglo-Saxon cemeteries: Berinsfield and Didcot.* Oxford: Oxford Archaeological Unit.

Härke H (2001) Population replacement or acculturation? An archaeological perspective on population and migrationin post-Roman Britain. *Celtic Englishes III Conference.*

Härke H (2002) Kings and warriors - Population and landscape from post-Roman to Norman Britain. In *The peopling of Britain: the shaping of a human landscape.*, pp. 145-175 [PA Slack and R Ward, editors]. Oxford: Oxford University Press.

Harpending HC (1994) Signature of ancient population-growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66**, 591-600.

Harpending H & Rogers A (2000) Genetic perspectives on human origins and differentiation. *Annual Review of Genomics and Human Genetics* **1**, 361-385.

Harpending H, Sherry S, Rogers A & Stoneking M (1993) The genetic structure of ancient human populations. *Current Anthropology* **34**, 483-496.

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR & Sherry ST (1998) Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1961-1967.

Harvati K (2003) The Neanderthal taxonomic position: models of intra- and inter-specific craniofacial variation. *Journal of Human Evolution* **44**, 107-132.

Hasegawa M, DiRienzo A, Kocher TD & Wilson AC (1993) Toward a More Accurate Time-Scale for the Human Mitochondrial- DNA Tree. *Journal of Molecular Evolution* **37**, 347-354.

Hasegawa M, Kishino H, Hayasaka K & Horai S (1990) Mitochondrial DNA evolution in primates: transition rate has been extremely low in the lemur. *Journal of Molecular Biology* **31**, 113-121.

Hauswirth WW, Dickel CD, Rowold DJ & Hauswirth MA (1994) Interpopulation and intrapopulation studies of ancient humans. *Experientia* **50**, 585-591.

Hawkes CFC (1967) An early Bronze age urn from Milton Northamptonshire. *Antiquaries Journal* **XLVII**, 198-208.

Hayden B (1972) Population control among hunter/gatheres. *World Archaeology* **4**, 206-221.

Hedges SB, Kumar S, Tamura K & Stoneking M (1991) Human origins and analysis of mitochondrial DNA sequences. *Science* **255**, 737-739.

Heillener K (1967) The population of Europe from the Black Death to the Eve of the Vital Revolution. In *The Cambridge economic history of Europe, Vol. 4: The economy of expanding Europe in the sixteenth and seventeenth centuries.* [EE Rich and CH Wilson, editors]. Cambridge: Cambridge University Press.

Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, Ward R & Sykes B (2001) mtDNA and the islands of the North Atlantic: Estimating the proportions of Norse and Gaelic ancestry. *American Journal of Human Genetics* **68**, 723-737.

Helgason A, Sigurðardóttir S, Gulcher J, Stefansson K & Ward R (2000a) Sampling saturation and the European mtDNA pool: Implicaitons for detecting relationships among populations. In *Archaeogenetics: DNA and the population prehistory of Europe.*, pp. 285-294 [C Renfrew and K Boyle, editors]. Cambridge: McDonald Institute for Archaeological Research, University of Cambridge.

Helgason A, Sigurðardóttir S, Gulcher JR, Ward R & Stefansson K (2000b) mtDNA and the origin of the Icelanders: Deciphering signals of recent population history. *American Journal of Human Genetics* **66**, 999-1016.

Hewitt G (1989) The subdivision of species by hibrid zones. In *Speciation and its consequences*, pp. 85-110 [D Otte and JA Endler, editors]. Sinauer: Sunderland, Mass.

Hewitt G (1999) Post-glacial re-colonization of Europea biota. *Biological Journal of the Linnean Society* **68**, 87-112.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.

Higuchi R (1984) DNA sequences from the quagga, an extint member of the horse family. *Nature* **312**, 282-284.

Hofreiter M, Jaenicke V, Serre D, von Haeseler A & Pääbo S (2001a) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research* **29**, 4793-4799.

Hofreiter M, Poinar HN, Spaulding WG, Bauer K, Martin PS, Possnert G & Pääbo S (2000) A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology* **9**, 1975-1984.

Hofreiter M, Serre D, Poinar HN, Kuch M & Pääbo S (2001b) Ancient DNA. *Nature Reviews Genetics* **2**, 353-359.

Holland MM, Fisher DL, Mitchell LG, Rodríguez WC, Canik JJ, Merril CR & Weedn VW (1993) Mitochondrial-DNA Sequence-Analysis of Human Skeletal Remains - Identification of Remains from the Vietnam-War. *Journal of Forensic Sciences* **38**, 542-553.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65-70.

Horai S (1989) Amplification from ancient human skeletal remains and their sequence analysis. *Proceedings of the Japan Academy* **65**, 229-233.

Horai S, Hayasaka K, Kondo R, Tsugane K & Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proceedings of the National Academy of Science USA* **92**, 532-536.

Hoss M (2000) Ancient DNA - Neanderthal population genetics. *Nature* **404**, 453-454.

Hoss M & Pääbo S (1993) DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Research* **21**, 3913-3914.

Housley RA, Gamble CS, Street M & Pettitt P (1997) Radiocarbon evidence for the lateglacial human recolonisation of northern Europe. *Proceedings of the Prehistoric Society* **63**, 25-54.

Howell N, Kubacka I & Mackey DA (1996) How rapidly does the human mitochondrial genome evolve? *American Journal of Human Genetics* **59**, 501-509.

Howell N & Mackey D (1997) mtDNA mutation rates - No need to panic - Reply. *American Journal of Human Genetics* **61**, 986-990.

Howell N & McCullough DA (1990) An example of Leber hereditary optic neuropathy not involving a mutation in the mitochondrial ND4 gene. *American Journal of Human Genetics* **47**, 629-634.

Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM & Herrnstadt C (2003) The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *American Journal of Human Genetics* **72**, 659-670.

Huhne J, Pfeiffer H, Waterkamp K & Brinkmann B (1999) Mitochondrial DNA in human hair shafts - existence of intra- individual differences? *International Journal of Legal Medicine* **112**, 172-175.

Hummel S & Herrmann B (1991) Y-chromosome-specific DNA amplified in ancient human bone. *Naturwissenschaften* **78**, 266-267.

Hummel S, Schmidt D, Kahle M & Herrmann B (2002) AB0 blood group genotyping of ancient DNA by PCR-RFLP. *International Journal of Legal Medicine* **116**, 327-333.

Hunter Blair P (1956) *An introduction to Anglo-Saxon England.* Cambridge: Cambridge University Press.

Hurlbert SH (1971) The non-concept of species diversity: a critique and alternative parameters. *Ecology* **52**, 577-586.

Hutchison CA-III, Newbold JE, Potter SS & Edgell MH (1974) Maternal inheritance of mammalian mitochondrial DNA. *Nature* **251**, 536-538.

Izagirre N & de la Rúa C (1999) An mtDNA analysis in ancient Basque populations: Implications for haplogroup V as a marker for a major Paleolithic expansion from southwestern Europe. *American Journal of Human Genetics* **65**, 199-207.

Jackson P (1995) Footloose in Archaeology. *Current Archaeology* **144**, 466-471.

Jackson P (1998) Footlose in Archaeology. *Current Archaeology* **156**, 456-457.

Jazin E, Soodyall H, Jalonen P, Lindholm E, Stoneking M & Gyllensten U (1998) Mitochondrial mutation rate revisited: hot spots and polymorphism. *Nature Genetics* **18**, 109-110.

Jeffreys AJ, Allen MJ, Hagelberg E & Sonnberg A (1992) Identification of the skeletal remains of Josef Mengele by DNA analysis. *Forensic Science International* **56**, 65-76.

Jehaes E, Pfeiffer H, Toprak K, Decorte R, Brinkmann B & Cassiman JJ (2001) Mitochondrial DNA analysis of the putative heart of Louis XVII, son of Louis XVI and Marie-Antoinette. *European Journal of Human Genetics* **9**, 185-190.

Jochim M (1983) Paleolithic cave art in ecological perspective. In *Hunther-gatherer economy in prehistory*, pp. 212-219 [G Bailey, editor]. Cambridge: Cambridge University Press.

Jochim M (1987) Late Plaistocene Refugia. In *The Pleistocene World - Regional Perspectives*, pp. 317-332 [O Soffer, editor]. New York: Plenum Press.

Johnson DA, Gautsch JW, Sportsman JR & Elder JH (1984) Improved technique utilizing non-fat dry milk for analysis of proteins and nucleic acid transferred to nitrocellulose. *Gene Analytical Techniques* **1**, 3-8.

Johnson H, Ricahrds PC, Long D & Graham CC (1993) *The geology of the northern North Sea.* London: HMSO.

Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC & Cavalli-Sforza LL (1983) Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. *Journal of Molecular Evolution* **19**, 255-271.

Jones M & Brown T (2000) Agricultural origins: the evidence of modern and ancient DNA. *Holocene* **10**, 769-776.

Jorde L & Bamshad M (2000) Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**, 1931a.

Jorde LB, Bamshad M, Watkins W, Zenger R, Fraley A, Krakowiak P, Carpenter K, Soodyall H, Jenkins T & Rogers A (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *American Journal of Human Genetics* **57**, 523-538.

Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Kere J & Harpending HC (1997) Microsatellite diversity and the demographic history of modern humans. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 3100-3103.

Juniper SK, Cambon MA, Lesongeur F & Barbier G (2001) Extraction and purification of DNA from organic rich subsurface sediments (ODP Leg 169S). *Marine Geology* **174**, 241-247.

Kaessmann H, Zollner S, Gustafsson AC, Wiebe V, Laan M, Lundeberg J, Uhlen M & Pääbo S (2002) Extensive linkage disequilibrium in small human populations in Eurasia. *American Journal of Human Genetics* **70**, 673-685.

Kahila Bar-Gal G, Khalaily H, Mader O, Ducos P & Kolska Horwitz L (2002) Ancient DNA evidence for the transition from wild to domestic status in Neolithic goats: A case study from the site of Abu Gosh, Israel. *Ancient Biomolecules* **4**, 9-18.

Kahn P & Gibbons A (1997) DNA from an extint human. *Science* **277**, 176-178.

Kalaydjieva L, Calafell F, Jobling MA, Angelicheva D, de Knijff P, Rosser ZH, Hurles ME, Underhill P, Tournev I, Marushiakova E & Popov V (2001) Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. *European Journal of Human Genetics* **9**, 97-104.

Kaldma K, Kivisild T, Calafell F, Kalaydjieva L, Parik J, Ferak V & Villems R (1998) Mitochondrial DNA lineages and the history of the Roms (Gypsies). NCBI (direct submission), AJ230488-AJ230614.

Keynes S (1997) The Vikings in England, c.790-1016. In *The Oxford ilustrated history of the Vikings*, pp. 48-82 [P Sawyer, editor]. Oxford: Oxford University Press.

Kiesslich J, Radacher M, Neuhuber F, H.J. M & Zeller KW (2002) On the use of nitrocellulose membranes for dialysis-mediated purification of ancient DNA from human bone and teeth extracts. *Ancient Biomolecules* **4**, 79-87.

Kimura B (1968) Evolutionary rate at the molecular level. *Nature* **217**, 624-626.

Kimura M & Crow JF (1963) The meassurement of effective population number. *Evolution* **17**, 279-288.

Kimura M & Crow JF (1964) The number of alleles that can be manteined in a finite population. *Genetics* **49**, 725-738.

King M-C & Wilson A (1975) Evolution at two levels in humans and chimpanzees. *Science* **188**, 107-166.

Kirk NJ, Wood RE & Goldstein M (2002) Skeletal identification using the frontal sinus region: A retrospective study of 39 cases. *Journal of Forensic Sciences* **47**, 318-323.

Kittles RA (1999) Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: Evidence for a male-specific bottleneck. *American Journal of Physical Anthropology* **108**, 381-399.

Kivisild T & Villems R (2000) Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**, 1931a.

Kolman CJ & Tuross N (2000) Ancient DNA analysis of human populations. *American Journal of Physical Anthropology* **111**, 5-23.

Kondo R, Matsuura ET & Chigusa SI (1992) Further observation of paternal transmission of Drosophila mitochondrial DNA by PCR selective amplification method. *Genetic Research* **59**, 81-84.

Kowaleski M (1984) The commercial dominance of a Medieval Provincial Oligarchy: Exeter in the late 14th century. *Medieval Studies* **46**, 355-384.

Kozlowski JK (1990) Northern Central Europe c. 18000 BP. In *The world at 18,000 BP. Volume 1: High latitudes*, pp. 204-227 [O Soffer and C Gamble, editors]. London: Unwin Hyman Inc.

Kreader C (1996) Relief of amplification inhibition in PCR with Bovine Serum Albumin or T4 Gene 32 Protein. *Applied Envronmental Microbiology.* **62**, 1102-1106.

Krings M, Capelli C, Tschentscher F, Geisert H, Meyer S, von Haeseler A, Grossschmidt K, Possnert G, Paunovic M & Pääbo S (2000) A view of Neandertal genetic diversity. *Nature Genetics* **26**, 144-146.

Krings M, Geisert H, Schmitz RW, Krainitzki H & Pääbo S (1999) DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 5581-5585.

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M & Pääbo S (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* **90**, 19-30.

Kumar S, Hedrick P, Dowling T & Stoneking M (2000) Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**, 1931a.

Kumar S, Tamura K, Jakobsen IB & Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244-1245.

Kumar SS, Tamura K & Nei M (1993) MEGA: molecular evolutionary genetics analysis. The Pennsylvania State University, University Park, PA 16802.

Kurosaki K, Matsushita T & Ueda S (1993) Individual identification from ancient human remains. *American Journal of Human Genetics* **53**, 638-643.

Kwok S & Higuchi R (1989) Avoiding false positives with PCR. *Nature* **339**, 237-238.

Lahermo P, Sajantila A, Sistonen P, Lukka M, Aula P, Peltonen L & Savontaus ML (1996) The genetic relationship between the Finns and the Finnish Saami (Lapps): Analysis of nuclear DNA and mtDNA. *American Journal of Human Genetics* **58**, 1309-1322.

Lalueza-Fox C, Shapiro B, Bover P, Alcover JA & Bertranpetit J (2002) Molecular phylogeny and evolution of the extinct bovid *Myotragus balearicus*. *Molecular Phylogenetics and evolution* **25**, 501-510.

Larruga JM, Diez F, Pinto FM, Flores C & González AM (2001) Mitochondrial DNA characterisation of European isolates: The Maragatos from Spain. *European Journal of Human Genetics* **9**, 708-716.

Leeds E (1945) *The settlement of the Anglo-Saxons - archaeologically considered.*

Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillen S & Vila C (2002) Ancient DNA evidence for Old World origin of New World dogs. *Science* **298**, 1613-1616.

Levett AE (1916) The Black Death on the States of the See of Winchester. In *Oxford studies in social and legal history, vol.5, The Black Death* [V Vinogradoff, editor]. Oxford: Clarendon Press.

Libert F, Cochaux P, Beckman G, Samson M, Aksenova M, Cao A, Czeizel A, Claustres M, de la Rúa C, Ferrari M, Ferrec C, Glover G, Grinde B, Guran S, Kucinskas V, Lavinha J, Mercier B, Ogur G, Peltonen L, Rosatelli C, Schwartz M, Spitsyn V, Timar L, Beckman L, Parmentier M & Vassart G (1998) The Delta ccr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. *Human Molecular Genetics* **7**, 399-406.

Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715.

Lindahl T (1997) Facts and artifacts of ancient DNA. *Cell* **90**, 1-3.

Logan FD (1983) *The Vikings in History*. London: Hutchinson & Co. Ltd.

Long D, Wickam-Jones CR & Ruckley NA (1986) A flint artefact from the northern North Sea. In *Studies in the Upper Paleolithic of Britain and Northwest Europe*, pp. 55-62 [DA Roe, editor]. Oxford: BAR international series.

Loyn HR (1984) *The Governance of Anglo-Saxon England 500-1087*. London.

Lucy S (1997) Burial Practice in Early Medieval Britain: constructing local identities. *Cultural Identities*.

Lucy S (2000) *The Anglo-Saxon way of death: burial rites in early England*. Sutton: Thrupp, Gloucestershire.

Lum JK & Cann RL (2000) mtDNA lineage analyses: origins and migrations of Micronesians and Polynesians. *American Journal of Physical Anthropology* **113**, 151-168.

Lundberg KS, Shoemaker DD, Adams MWW, Short JM, Sorge JA & Mathur EJ (1991) High-Fidelity Amplification Using a Thermostable DNA-Polymerase Isolated from Pyrococcus-Furiosus. *Gene* **108**, 1-6.

Lutz S, Weisser HJ, Heizmann J & Pollak S (1996) mtDNA as a tool for identification of human remains - Identification using mtDNA. *International Journal of Legal Medicine* **109**, 205-209.

Lutz S, Weisser HJ, Heizmann J & Pollak S (1999) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany (vol 111, pg 67, 1998). *International Journal of Legal Medicine* **112**, 145-150.

Lutz S, Weisser H-J, Heizmann J & Pollak S (2000a) Mitochondrial heteroplasmy among maternally related individuals. *International Journal of Legal Medicine* **113**, 155-161.

Lutz S, Wittig H, Weisser HJ, Heizmann J, Junge A, Dimo-Simonin N, Parson W, Edelmann J, Anslinger K, Jung S & Augustin C (2000b) Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? *Forensic Science International* **113**, 97-101.

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B & Torroni A (1999a) The emerging tree of west Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs. *American Journal of Human Genetics* **64**, 232-249.

Macaulay V, Richards M & Sykes B (1999b) Mitochondrial DNA recombination - no need to panic. *Proceedings of the Royal Society of London Series B-Biological Sciences* **266**, 2037-2039.

Macaulay VA, Richards MB, Forster P, Bendall KE, Watson E, Sykes B & Bandelt HJ (1997) mtDNA mutation rates - No need to panic. *American Journal of Human Genetics* **61**, 983-986.

Machugh D, Edwards CJ, Bailey JF, Bancroft DR & Bradley D (2000) The extraction and anlyses of ancient DNA from bone and teeth: a survey from current methodologies. *Ancient Biomolecules* **3**, 81-102.

Magurran AE (1988) *Ecological diversity and its meassure*. Princeton.

Malhi RS & Smith DG (2002) Brief communication: Haplogroup X confirmed in prehistoric North America. *American Journal of Physical Anthropology* **119**, 84-86.

Mall G, Hubig M, Buttner A, Kuznik J, Penning R & Graw M (2001) Sex determination and estimation of stature from the longbones of the arm. *Forensic Science International* **117**, 23-30.

Malyarchuk B & Derenko M (2001) Mutational hotspots and hypervariable nucleotide motifs in the non-coding human mtDNA control region. *American Journal of Human Genetics* **69**, 1220.

Markina SP, Belyaeva OV, Slominsky PA, Bebyakova NA & Limborska SA (2000) Mitochondrial DNA diversity in eastern Slavonic populations. NCBI (direct submission), AY005827-AY005866.

Marlow M (1992) The human remains: the population. In *An Anglo-Saxon cemetery at Norton, Cleveland.*, pp. 107-118: Council for British Archaeology.

Marota I, Basile C, Ubaldi M & Rollo F (2002) DNA decay rate in papyri and human remains from Egyptian archaeological sites. *American Journal of Physical Anthropology* **117**, 310-318.

Martinson JJ, Chapman NH, Rees DC, Liu YT & Clegg JB (1997) Global distribution of the CCR5 gene 32-basepair deletion. *Nature Genetics* **16**, 100-103.

Mascie-Taylor CGN & Lasker GW (1996) Further notes on possible changes in the geographic distribution of ABO and Rh blood groups in Great Britain. *Human Biology* **68**, 473-478.

Mastana SS & Sokol RJ (1998) Genetic variation in the East Midlands. *Annals of Human Biology* **25**, 43-68.

Matheson CD & Loy TH (2001) Genetic Sex Identification of 9400-year-old Human Skull Samples from Çayönü Tepesi, Turkey. *Journal of Archaeological Science* **28**, 569-575.

McEvedy C (1988) The bubonic plague. *Scientific American* **258**, 118-123.

McEvedy C & Jones R (1978) *Atlas of world population history.* London: Allen Lane.

Mellars P (1974) The Paleolithic and Mesolithic. In *British Prehistory*, pp. 41-99 [C Renfrew, editor]. New Jersey: Noyes Press, Park Ridge.

Mellars P (1989) Major issues in the emergence of modern humans. *Current Anthropology* **30**, 349-385.

Mellars P (1992) Archaeology and the population-dispersal hypothesis of modern human origins in Europe. *Philosophical Transactions of the Royal Society of London. Series B.* **337**, 225-234.

Mellars P (1993) Archaeology and modern human origins in Europe. *Proceedings of the British Academy* **82**, 1-35.

Melton T, Wilson M, Batzer M & Stoneking M (1997) Extent of heterogeneity in mitochondrial DNA of European populations. *Journal of Forensic Sciences* **42**, 437-446.

Menozzi P, Piazza A & Cavalli-Sforza L (1978) Synthetic maps of human gene frequencies in Europeans. *Science* **201**, 786-792.

Merriweather DA (2000) Ancient DNA and human migrations. *5th International Ancient DNA Conference.*

Merriwether DA, Rothhammer F & Ferrell RE (1994) Genetic variation in the New-World - Ancient teeth, bone, and tissue as sources of DNA. *Experientia* **50**, 592-601.

Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, García LF, Vélez ID, Bravo ML, Munera JG, Bedoya G, Bortolini MC & Ruiz-Linares A (2000) Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: Pre- and post-Columbian patterns of gene flow in South America. *American Journal of Human Genetics* **67**, 1277-1286.

Meyer S, Weiss G & von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* **152**, 1103-1110.

Miller DN (2001) Evaluation of gel filtration resins for the removal of PCR- inhibitory substances from soils and sediments. *Journal of Microbiological Methods* **44**, 49-58.

Miller KWP, Dawson JL & Hagelberg E (1996) A concordance of nucleotide substitutions in the first and second hypervariable segments of the human mtDNA control region. *International Journal of Legal Medicine* **109**, 107-113.

Mirazón Lahr M, Foley RA & Pinhasi R (2000) Expected regional patterns of Mesolithic-Neolithic human population admixture in Europe based on archaeological evidence. In *Archaeogenetics: DNA and the popualtion prehistory of Europe*, pp. 81-88 [C Renfrew and AJ Boyce, editors]. Cambridge: McDonald Institute for Archaeological Research, University of Cambridge.

Montiel R, Malgosa A & Francalacci P (2001) Authenticating ancient human mitochondrial DNA. *Human Biology* **73**, 689-713.

Montiel R, Malgosa A & Subirà E (1997) Overcoming PCR inhibitors in ancient DNA extracts from teeth. *Ancient Biomolecules* **1**, 221-225.

Montiel-Duarte R (2000) Estudio Diacronico de la variabilidad del ADN mitocondrial en la poblacion Catalana., Universidad Autonoma de Barcelona.

Moreira D (1998) Efficient removal of PCR inhibitors using agarose-embedded DNA preparations. *Nucleic Acids Research* **26**, 3309-3310.

Moretti T, Koons B & Budowle B (1998) Enhancement of PCR amplification yield and specificity using AmpliTaq Gold ™ polymerase. *BioTechniques* **25**, 716-722.

Morris AAM & Lightowlers RN (2000) Can paternal mtDNA be inherited? *Lancet* **355**, 1290-1291.

Mullis K & Faloona F (1987) Specific synthesis of DNA in vitro via polymerase-catalised chain reaction. *Methodological Enzymology* **155**, 335-350.

Mumm S, Whyte MP, Thakker RV, Buetow KH & Schlessinger D (1997) mtDNA analysis shows common ancestry in two kindreds with X-linked recessive hypoparathyroidism and reveals a heteroplasmic silent mutation. *American Journal of Human Genetics* **60**, 153-159.

Myres JNL (1986) *The Anglo-Saxon settlements*. Oxford.

Nei M (1975) *Molecular population genetics and evolution*. Amsterdam: North-Holland Publishing Co.

Nei M (1987) *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nei M & Roychoudhury AK (1988) *Human polymorphic genes: world distribution*. New York: Oxford University Press.

Noonan T (1997) Scandinavians in European Russia. In *The Oxford illustrated history of the Vikings*, pp. 134-155 [P Sawyer, editor]. Oxford: Oxford University Press.

Norborg M (1998) On the probability of Neanderthal ancestry. *American Journal of Human Genetics* **63**, 1237-1240.

Oota H, Kurosaki K, Pookajorn S, Ishida T & Ueda S (2001a) Genetic study of the Paleolithic and Neolithic Southeast Asians. *Human Biology* **73**, 225-231.

Oota H, Saitou N, Matsushita T & Ueda S (1995) A genetic study of 2,000-year-old human remains from Japan using mitochondrial DNA sequences. *American Journal of Physical Anthropology* **98**, 133-145.

Oota H, Saitou N, Matsushita T & Ueda S (1999) Molecular genetic analysis of remains of a 2,000-year-old human population in China - and its relevance for the origin of the modern Japanese population. *American Journal of Human Genetics* **64**, 250-258.

Oota H, Settheetham-Ishida W., Tiwawech D, Ishida T & Stoneking M (2001b) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nature Genetics* **29**, 20-21.

Orlando L, Bonjean D, Bocherens H, Thenot A, Argant A, Otte M & Hanni C (2002) Ancient DNA and the population genetics of cave bears (*Ursus spelaeus*) through space and time. *Molecular Biology and Evolution* **19**, 1920-1933.

O'Rourke DH, Carlyle SW & Parr RL (1996) Ancient DNA: methods, progress and perspectives. *American Journal of Human Biology* **8**, 557-571.

Otte M (1976) L'Aurignacien en Belgique. In *L-Arignacien en Europe*, pp. 144-163. Nice: Actes du IXe Congres de l'Union International des Sciences Prehistoriques et Prehistoriques.

Otte M & Keeley LH (1990) The impact on regionalism in Paleolithic studies. *Current Anthropology* **31**, 577-582.

Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K & Goodwin W (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* **404**, 490-493.

Owen O (1999) *The sea road. A Viking voyage through Scotland*: Canongate books with Historic Scotland.

Pääbo S (1985) Molecular clonning of ancient Egyptian mummy DNA. *Nature* **314**, 644-645.

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

Pääbo S (1996) Mutational hot spots in the mitochondrial microcosm. *American Journal of Human Genetics* **59**, 493-496.

Pääbo S, Gifford JA & Wilson AC (1988) Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids Research* **16**, 9775-9787.

Pääbo S, Higuchi R & Wilson A (1989a) Ancient DNA and the polymerase chain reaction. *The Journal of Biological Chemistry* **264**, 9709-9712.

Pääbo S, Irwin DM & Wilson A (1989b) DNA damage promotes jumping between templates during enzymatic amplification. *The Journal of Biological Chemistry* **265**, 4718-4721.

Pares JM & Pérez-González A (1995) Paleomagnetic age for hominid fossils at Atapuerca archaeological site, Spain. *Science* **269**, 830-832.

Parfitt K (1995) The Buckland Anglo-Saxon cemetery. *Current Archaeology* **144**, 459-464.

Parr RL, Carlyle SW & O'Rourke DH (1996) Ancient DNA analysis of Fremont Amerindians of the Great Salt Lake Wetlands. *American Journal of Physical Anthropology* **99**, 507-518.

Parson W, Parsons TJ, R. S & Holland MM (1998) Population data for 101 Austrian Caucasian mitochondrial DNA D-loop sequences: application of mtDNA sequence analysis to a forensic case. *International Journal of Legal Medicine* **111**, 124-132.

Parsons TJ & Holland MM (1998) Mitochondrial mutation rate revisited: hot spots and polymorphism - Response. *Nature Genetics* **18**, 110-110.

Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P & Holland MM (1997) A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics* **15**, 363-368.

Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Borresen-Dale AL & Underhill PA (2002) Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. *European Journal of Human Genetics* **10**, 521-529.

Paterson R (2002) Yersinia seeks pardon for Black Death. *Lancet Infectious Diseases* **2**, 323-323.

Penn K (2000) Anglo-Scandinavian Norwich: an archaeological desk-top survey. Norfolk: Report No. 488. North Sea Viking Legacy Project. Norfolk Archaeological Unit.

Perkins DJ (1969) Fauna of Catal Huyuk: evidence for early cattle domestication in Anatolia. *Science* **164**, 177-179.

Petit RJ, El Mousadik A & Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**, 844-855.

Pfeiffer H, Huhne J, Seitz B & Brinkmann B (1999) Influence of soil storage and exposure period on DNA recovery from teeth. *International Journal of Legal Medicine* **112**, 142-144.

Piercy R, Sullivan KM, Bensom N & Gill P (1993) The application of mitocondrial DNA typing to the estudy of white Caucasian genetic identification. *International Journal of Legal Medicine* **106**, 85-90.

Pinhasi R, Foley RA & Mirazón Lahr M (2000) Spatial and temporal patterns in the Mesolithic-Neolithic archaeological record of Europe. In *Archaeogenetics: DNA and the popualtion prehistory of Europe*, pp. 45-56 [C Renfrew and K Boyle, editors]. Cambridge: McDonald Institute for Archaeological Research, University of Cambridge.

Pinto F, González AM, Hernández M, Larruga JM & Cabrera VM (1996) Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Annals of Human Genetics* **60**, 321-330.

Pitts MW & Jacobi RM (1979) Some aspects of change in flaked stone industries of the Mesolithic and Neolithic in southern Britain. *Journal of Archaeological Science* **6**, 163-177.

Poinar HN (1976) Aspartic acid racemization in dentin as a measure of ageing. *Nature* **262**, 279-281.

Popescu E (in prep.) Exacavation at Norwich Castle 1987-98. Norfolk: NAU.

Pretty IA & Sweet D (2001) A look at forensic dentistry - Part 1: The role of teeth in the determination of human identity. *British Dental Journal* **190**, 359-366.

Previdere C, Micheletti P, Perossa R, Grignani P & Fattorini P (2002) Molecular characterisation of the nucleic acids recovered from aged forensic samples. *International Journal of Legal Medicine* **116**, 334-339.

Primorac D, Andelinovic S, DefinisGojanovic M, Drmic I, Rezic B, Baden MM, Kennedy MA, Schanfield MS, Skakel SB & Lee HC (1996) Identification of war victims from mass graves in Croatia, Bosnia, and Herzegovina by the use of standard forensic methods and DNA typing. *Journal of Forensic Sciences* **41**, 891-894.

Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W & Pääbo S (1994) Mitochondrial-DNA sequences from Switzerland reveal striking homogeneity of European populations. *Biological Chemistry Hoppe-Seyler* **375**, 837-840.

Pusch C & Scholtz M (1997) DNA extraction from ancient human bones via enzymatic treatment. *Elesevier Trends Journals Technical Tips on-line.*

Rameckers J, Hummel S & Herrmann B (1997) How many cycles does a PCR need? Determination of cycle numbers depending on the number of targets and reaction efficiency factor. *Naturwissenschaften* **84**, 259-262.

Raoult D, Aboudharam G, Crubezy E, Larrouy G, Ludes B & Drancourt M (2000) Molecular identification by "suicide PCR" of Yersinia pestis as the agent of Medieval Black Death. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 12800-12803.

Raoult D & Drancourt M (2002) Cause of Black Death. *Lancet Infectious Diseases* **2**, 459-459.

Reese KM (2002) Black Death not bubonic plague. *Chemical & Engineering News* **80**, 56-56.

Reese RL, Mawk EJ, Derr JN, Hyman M, Rowe MW & Davis SK (1996) Ancient DNA in Texas rock paintings. *Archaeological Chemistry* **625**, 378-390.

Reidla M, Metspalu E, Kivisild T, Kaldma K & Villems R (1999) Mitochondrial DNA lineages in Georgia. NCBI (direct submission), AJ389197-AJ389320.

Relethford JH (2001) Absence of regional affinities of Neandertal DNA with living humans does not reject multiregional evolution. *American Journal of Physical Anthropology* **115**, 95-98.

Remington CL (1968) Suture-zones of hybrid interaction between recently joined biotas. *Evolutionary Biology* **2**, 321-428.

Rendine S, Piazza A & Cavalli-Sforza LL (1986) Simultion and separation by pricipal components of multiple demic expansions in Europe. *American Naturalist* **128**, 681-706.

Renfrew C (1987) *Archaeology and language: the puzzle of Indo-European origins*. London.

Renfrew C (2000) Introduction. In *Archaeogenetics: DNA and the population prehistory of Europe.* [C Renfrew and K Boyle, editors]. Cambridge: McDonald Institute for Archaeological Research, University of Cambridge.

Reynier P, May-Panloup P, Chretien MF, Morgan CJ, Jean M, Savagner F, Barriere P & Malthiery Y (2001) Mitochondrial DNA content affects the fertilizability of human oocytes. *Molecular Human Reproduction* **7**, 425-429.

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

Richards M & Macaulay V (2000) Genetic data and the colonization of Europe: genealogies and founders. In *Archaeogenetics: DNA and the popualtion prehistory of Europe*, pp. 139-150 [C Renfrew and K Boyle, editors]. Cambridge: McDonald Institute Monographs.

Richards M, Macaulay V, Bandelt H-J & Sykes B (1998a) Phylogeography of mitochondrial DNA in western Europe. *Annals of Human Genetics* **62**, 241-260.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Golge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Norby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozzari R, Torroni A & Bandelt HJ (2000) Tracing European founder lineages in the near eastern mtDNA pool. *American Journal of Human Genetics* **67**, 1251-1276.

Richards M, Macaulay V, Sykes B, Pettitt P, Hedges R, Forster P & Bandelt H-J (1997) Reply to Cavalli-Sforza and Minch. *American Journal of Human Genetics* **61**, 251-254.

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

Richards M, Oppenheimer S & Sykes B (1998b) MtDNA suggests Polynesian origins in eastern Indonesia. *American Journal of Human Genetics* **63**, 1234-1236.

Richards M, Smalley K, Sykes B & Hedges R (1993) Archaeology and genetics: analysing DNA from skeletal remains. *World Archaeology* **25**, 18-28.

Richards MB, Sykes B & Hedges R (1995) Authenticating DNA extracted from ancient skeletal remains. *Journal of Archaeological Science* **22**, 291-299.

Rickards O, Martinez-Labarga C, Favaro M, Frezza D & Mallegni F (2001) DNA analyses of the remains of the Prince Branciforte Barresi family. *International Journal of Legal Medicine* **114**, 141-146.

Rightmire GP (1984) *Homo sapiens* in the Sub-Saharan Africa. In *The origins of modern humans: A world survey of the fossil record*, pp. 295-326 [FH Smith and F Spencer, editors]. New York: Liss.

Ritchie A (1995) *Prehistoric Orkney*. London: Batsford, Historic Scotland.

Ritte U, Neufeld E, Prager EM, Gross M, Hakim I, Khatib A & Bonne-Tamir B (1993) Mitochondrial DNA affinity of several Jewish communities. *Human Biology* **65**, 359-385.

Robin ED & Wong R (1988). Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *Journal of Cellular Physiology* **136**, 507-513.

Rogers AR & Harpending H (1992) Population-growth makes waves in the distribution of pairwise genetic distances. *Molecular Biology and Evolution* **9**, 552-569.

Röhl A (1997) Network 2.0: a program package for calculating phylogenetic networks. *Available at http://www.fluxus-engineering.com/.*

Röhl A, Brinkmann B, Forster L & Forster P (2001) An annotated mtDNA database. *International Journal of Legal Medicine* **115**, 29-39.

Rozas J & Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174-175.

Ruhlen M (1987) *A guide to the world's languages.* London: Edward Arnold.

Ruitberg CM, Reeder DJ & Butler JM (2001) STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Research* **29**, 320-322.

Russell JC (1948) *British Medieval population*. Albuquerque: University of Mexico.

Sajantila A & Budowle B (1991) Identification of Individuals with DNA Testing. *Annals of Medicine* **23**, 637-642.

Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjaerg L, Geddedahl T, Isseltarver L, Dirienzo A & Pääbo S (1995) Genes and Languages in Europe - an Analysis of Mitochondrial Lineages. *Genome Research* **5**, 42-52.

Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C & Pääbo S (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 12035-12039.

Salas A, Lareu V, Calafell F, Bertranpetit J & Carracedo A (2000) mtDNA hypervariable region II (HVII) sequences in human evolution studies. *European Journal of Human Genetics* **8**, 964-974.

Sambrook J, Fritsch EF & Maniatis T (1989) *Molecular Clonning: a laboratory manual*. New York: Cold Spring Harbour Press.

Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, Saragosti S, Lapoumeroulie C, Cognaux J, Forceille C, Muyldermans G, Verhofstede C, Burtonboy G, Georges M, Imai T, Rana S, Yi Y, Smyth RJ, Collman RG, Doms RW, Vassart G & Parmentier M (1996) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722-725.

Sanders HL (1968) Marine benthic diversity: a comparative study. *American Naturalist* **102**, 243-282.

Sarkar G & Sommer SS (1993) Removal of DNA contamination in polymerase chain reaction reagents by ultraviolet irradiation. *Methods in Enzymology* **218**, 381-389.

Sawyer P (1997) The age of theVikings, and before. In *The Oxford illustrated history of the Vikings*, pp. 1-18 [P Sawyer, editor]. Oxford: Oxford University Press.

Schmerer W, Hummel S & Herrmann B (1999) Optimized DNA extraction to improve reproducibility of short tandem repeat genotyping with highly degraded DNA as target . *Electrophoresis* **20**, 1712-1716.

Schmidt TM, Hummel S & Herrmann B (1995) Evidence of contamination in PCR laboratory disposables. *Naturwissenschaften* **82**, 423-431.

Schneider S, Kueffen J, Roessli D & Excoffier L (2001) Arlequin: A software for population genetics data analysis. Ver 2.100. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM & Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *American Journal of Human Genetics* **46**, 613-623.

Schurr TG, Sukernik RI, Starikovskaya EB & Wallace DC (1999) Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea-Bering Sea region during the Neolithic. *American Journal of Physical Anthropology* **108**, 1-39.

Schwartz M & Vissing J (2002) Paternal inheritance of mitochondrial DNA. *New England Journal of Medicine* **347**, 576-580.

Schwartz TR, Schwartz EA, Mieszerski L, McNally L & Kobilinsky L (1991) Characterization of deoxiribonucleic acid (DNA) obtained from teeth subjected to various environmental conditions. *Journal of Forensic Science*, 979-990.

Seielstad M, Minch E & Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. *Nature Genetics* **20**, 278-280.

Sequencher: Software for Macintosh, version 3.0. Ann Arbor MI, USA: Gene Codes Corporation.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL & Underhill PA (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: A Y chromosome perspective. *Science* **290**, 1155-1159.

Sherlock SJ & Welch MG (1992) *An Anglo-Saxon cemetery at Norton, Cleveland.*: Council for British Archaeology.

Shibutani S, Takeshita M & Grollman AP (1991) Insertion of Specific Bases During DNA-Synthesis Past the Oxidation-Damaged Base 8-Oxodg. *Nature* **349**, 431-434.

Sigurðardóttir S, Helgason A, Gulcher JR, Stefansson K & Donnelly P (2000) The mutation rate in the human mtDNA control region. *American Journal of Human Genetics* **66**, 1599-1609.

Simberloff DS (1972) Properties of the rarefaction diversity measurement. *American Naturalist* **106**, 414-418.

Simon J & Rigby V (1997) *Britain and the Celtic Iron Age*. London: British Museum.

Simoni L, Calafell F, Bertranpetit J & Barbujani G (2000a) Spatial variation of mtDNA hypervariable region I among European populations. In *Archaeogenetics: DNA and the population prehistory of Europe*, pp. 131-138 [C Renfrew and K Boyle, editors]. Cambridge: McDonald Institute for Archaeological Research, University of Cambridge.

Simoni L, Calafell F, Pettener D, Bertranpetit J & Barbujani G (2000b) Geographic patterns of mtDNA diversity in Europe. *American Journal of Human Genetics* **66**, 262-278.

Simoni L, Calafell F, Pettener D, Bertranpetit J & Barbujani G (2000c) Reconstruction of prehistory on the basis of genetic data. *American Journal of Human Genetics* **66**, 1177-1179.

Slatkin M & Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555-562.

Smith BC, Fisher DL, Weedn VW, Warnock GR & Holland MM (1993) A systematic approach to the sampling of dental DNA. *Journal of Forensic Science*, 1194-1209.

Smith LS, Lewis TL & Matsui SM (1995) Increased yield of small DNA fragments purified by silica binding. *BioTechniques* **18**, 970-975.

Soffer O (1985) *The Upper Paleolithic of the Central Russian Plain*. Orlando: Academic Press.

Soffer O (1987) Upper Paleolithic Connubia, refugia and the archaeological record from Eastern Europe. In *The Pleistocene World - Regional Perspectives*, pp. 333-348 [O Soffer, editor]. New York: Plenum Press.

Soffer O (1990) The Russian Plain at the Last Glacial Maximum. In *The world at 18,000 BP. Volume 1: High latitudes*, pp. 228-252 [O Soffer and C Gamble, editors]. London: Unwin Hyman Inc.

Soffer O & Gamble C (1990) Introduction. In *The world at 18,000 BP. Volume 1: High latitudes.* [O Soffer and C Gamble, editors]. London: Unwin Hyman Inc.

Sokal RR & Menozzi P (1982) Spatial autocorrelations of HLA frequencies in Europe support demic diffusion of early farmers. *American Naturalist* **119**, 1-17.

Sokal RR & Oden NL (1978) Spatial autocorrelation in Biology. *Biological Journal of the Linnean Society* **10**, 199-249.

Sokal RR, Oden NL & Wilson C (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* **351**, 143-145.

Spuhler JN (1988) Evolution of mitochondria in monkeys, apes and humans. *Yearbook of Physical Anthropology* **31**, 15-48.

St John J, Sakkas D, Dimitriadi C, Barnes A, Maclin V, Ramey J, Barratt C & De Jonge C (2000) Failure of elimination of paternal mitochondrial DNA in abnormal embryos. *Lancet* **355**, 200-200.

Stenico M, Nigro L, Bertorelle G, Calafell F, Capitanio M, Corrain C & Barbujani G (1996) High mitochondrial sequence diversity in linguistic isolates of the Alps. *American Journal of Human Genetics* **59**, 1363-1375.

Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M, Winkler C, Huttley GA, Allikmets R, Schriml L, Gerrard B, Malasky M, Ramos MD, Morlot S, Tzetis M, Oddoux C, di Giovine FS, Nasioulas G, Chandler D, Aseev M, Hanson M, Kalaydjieva L, Glavac D, Gasparini P, Kanavakis E, Claustres M, Kambouris M, Ostrer H, Duff G, Baranov V, Sibul H, Metspalu A, Goldman D, Martin N, Duffy D, Schmidtke J, Estivill X, O'Brien SJ & Dean M (1998) Dating the origin of the CCR5-Delta 32 AIDS-resistance allele by the coalescence of haplotypes. *American Journal of Human Genetics* **62**, 1507-1515.

Stone AC & Stoneking M (1993) Ancient DNA from a Pre-Columbian Amerindian Population. *American Journal of Physical Anthropology* **92**, 463-471.

Stone AC & Stoneking M (1998) mtDNA analysis of a prehistoric Oneota population: Implications for the peopling of the new world. *American Journal of Human Genetics* **62**, 1153-1170.

Stoneking M (1998) Women on the move. *Nature Genetics* **20**, 219-220.

Stoneking M, Hedgecock D, Higuchi R, Vigilant L & Elrich H (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *American Journal of Human Genetics* **48**, 370-382.

Stoneking M, Sherry ST, Redd AJ & Vigilant L (1992) New approaches to dating suggest a recent age for the human mtDNA ancestor. *Philosophical Transactions of the Royal Society of London. Series B.* **1280**, 167-175.

Strauss LG (1990) The Last Glacial Maximum in Cantabrian Spain: the Salutrean. In *The world at 18,000 BP. Volume 1: High latitudes*, pp. 89-108 [O Soffer and C Gamble, editors]. London: Unwin Hyman Inc.

Strauss LG (1995) The Upper Paleolithic in Europe: an overview. *Evolutionary Anthropology* **4**, 4-16.

Strauss LG, Clark GA, Altuna J & Ortea JA (1980) Ice Age subsistence in northern Spain. *Scientific American* **242**, 142-152.

Stringer C (1993) Exploring modern human origins - Progress and prospects. *Current Anthropology* **34**, 176-176.

Stringer C (2002) Modern human origins: progress and prospects. *Philosophical Transactions of the Royal Society of London. Series B.* **357**, 563-579.

Stringer C, Hublin JJ & Vandermeersch B (1984) The origin of anatomically modern humans in western Europe. In *The origin of modern humans: A world survey of the fossil evidence*, pp. 51-136 [FH Smith and F Spencer, editors]. New York: Liss.

Stringer CB & Andrews P (1988) Genetic and fossil evidence for the origin of modern humans. *Science* **239**, 1263-1265.

Sullivan K, Hopgood R & Gill P (1992) Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *International Journal of Legal Medicine* **105**, 83-86.

Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C & Schatten G (1999) Ubiquitin tag for sperm mitochondria. *Nature* **402**, 371-372.

Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C & Schatten G (2000) Ubiquitinated Sperm Mitochondria, Selective Proteolysis, and the Regulation of Mitochondrial Inheritance in Mammalian Embryos. *Biology of Reproduction* **63**, 582-590.

Sweet D, Lorente M, Valenzuela A, Lorente JA & Alvarez JC (1996) Increasing DNA extraction yield from saliva stains with a modified Chelex method. *Forensic Science International* **83**, 167-177.

Sykes B (1999) The molecular genetics of European ancestry. *Philosophical Transactions of the Royal Society of London. Series B.* **354**, 131-138.

Sykes B (2001) *The seven daughters of Eve.* London: Bantam.

Sykes B, Leiboff A, Lowbeer J, Tetzner S & Richards M (1995) The Origins of the Polynesians - an Interpretation from Mitochondrial Lineage Analysis. *American Journal of Human Genetics* **57**, 1463-1475.

Taberlet P, Fumagalli L, Wust-Saucy AG & Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**, 453-464.

Tajima F (1983) Evolutionary relationships of DNA sequence in finite size populations. *Genetics* **105**, 437-460.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.

Tambiah SJ (1973) From varna to caste through mixed unions. In *The character of kinship.* [J Goody, editor]. London: Cambridge University Press.

Tamura K & Nei M (1993) Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial-DNA in Humans and Chimpanzees. *Molecular Biology and Evolution* **10**, 512-526.

Tatersall I, Delson I & Van Couvering J (1988) *Encyclopaedia of human evolution and prehistory.* New York: Garland.

Tchou J, Kasai H, Shibutani S, Chung MH, Laval J, Grollman AP & Nishimura S (1991) 8-Oxoguanine (8-Hydroxyguanine) DNA Glycosylase and Its Substrate-Specificity. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 4690-4694.

Thomas M (1997) DNA extraction using the silica method. *Personal Communication.*

Thomas MG & Moore LJ (1996) Preparation of bone samples for DNA extraction: a nuts and bolts approach. *BioTechniques* **22**, 402.

Thomas MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scozzari R, Gratrix F, Tarekegn A, Wilson JF, Capelli C, Bradman N & Goldstein DB (2002) Founding mothers of Jewish communities: Geographically separated Jewish groups were independently founded by very few female ancestors. *American Journal of Human Genetics* **70**, 1411-1420.

Thomas RH, Schaffner W, Wilson AC & Pääbo S (1989) DNA phylogeny of the extinct marsupial wolf. *Nature* **340**, 465-467.

Thommessen T (1996) The early settlement of Northern Norway. In *The earliest settlement in Scandinavia and its relationship with neighbouring areas*, pp. 235-241 [L Larsen, editor]. Stockholm: Almquist & Wiksell International.

Thorne AG & Wolpoff MH (1981) Regional continuity in Australasian Pleistocene hominid evolution. *American Journal of Physical Anthropology* **55**, 337-349.

Thorpe IJ (1996) *The origins of agriculture in Europe*. London: Routldge.

Thwaites G, Taviner M & Gant V (1997) The English sweating sickness 1485-1551. *The New England Journal of Medicine* **336**, 580-582.

Tikochinski Y, Ritte U, Gross SR, Prager EM & Wilson AC (1991) mtDNA polymorphism in two communities of Jews. *American Journal of Human Genetics* **48**, 129-136.

Tofanelli S, Dell'Amico MC, Nencioni L, Paoli G & Borgognini Tarli SM (1999) Recovering ancient DNA by streptavidin-coated magnetic beads and biotinylated oligonucleotides. *Ancient Biomolecules* **2**, 307-320.

Torroni A, Bandelt H, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus M, BonneTamir B & Scozzari R (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *American Journal of Human Genetics* **62**, 1137-1152.

Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E, Parik JR, Tolk HV, Tambets K, Forster P, Karger B, Francalacci P, Rudan P, Janicijevic B, Rickards O, Savontaus ML, Huoponen K, Laitinen V, Koivumaki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Benerecetti AS, Semino O & Scozzari R (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *American Journal of Human Genetics* **69**, 844-852.

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML & Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**, 1835-1850.

Torroni A, Lott M, Cabell MF, Chen YS, Lavergne L & Wallace DC (1994) mtDNA and the origin of Caucasians: identification of Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *American Journal of Human Genetics* **55**, 760-776.

Torroni A, Richards M, Macaulay V, Forster P, Villems R, Norby S, Savontaus ML, Huoponen K, Scozzari R & Bandelt HJ (2000) mtDNA haplogroups and frequency patterns in Europe. *American Journal of Human Genetics* **66**, 1173-1177.

Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG & Wallace DC (1993) Mtdna variation of aboriginal Siberians reveals distinct genetic affinities with Native-Americans. *American Journal of Human Genetics* **53**, 591-608.

Torroni A & Wallace DC (1994) Mitochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. *Journal of Bioenergetics and Biomembranes* **26**, 261-271.

Torroni A, Yang C-C, Szathamary E, Williams R, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM & Wallace DC (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene popaultions were found by two independent migrations. *Genetics* **130**, 153-162.

Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P, Chamberlain AT, Sykes B & Bradley DG (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**, 1088-1091.

Ucko PJ & Dimbleby GW (1969) *The domestication and exploitation of plants and animals*. London.

Uhl J (2001) The secrets of the death - the mistery of the Black Death: Thirteen/WNET New York in association with Channel 4.

Underhill PA, Shen PD, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL & Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. *Nature Genetics* **26**, 358-361.

Varesi L, Memmi M, Cristofari MC, Mameli GE, Calo CM & Vona G (2000) Mitochondrial control-region sequence variation in the Corsican population, France. *American Journal of Human Biology* **12**, 339-351.

Vawter L & Brown WM (1986) Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. *Science* **234**, 194-196.

Vendramin GG, Degen B, Petit RJ, Anzidei M, Madaghiele A & Ziegenhagen B (1999) High level of variation at Abies alba chloroplast microsatellite loci in Europe. *Molecular Ecology* **8**, 1117-1126.

Verhardt LBM (1995) Fishing in the Mesolithic. The North Sea: a submerged Mesolithic landscape. In *Man & sea in the Mesolithic - Coastal settlement above and below present sea level. Proceedings of the International Symposium, Kalundborg, Denmark 1993*, pp. 291-296 [A Fischer, editor]. Oxford: Oxbow Monographs.

Vigilant L, Pennington R, Harpending H, Kocher TD & Wilson A (1989) Mitochondrial DNA sequences in single hairs from southern African population. *Proceedings of the National Academy of Sciences, USA* **86**, 9350-9354.

Vigilant L, Stoneking M, Harpending H, Hawkes K & Wilson AC (1991a) African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503-1507.

Vigilant L, Wilson AC & Harpending H (1991b) Evolution of human mitochondrial DNA in African populations. *American Journal of Human Genetics* **49**, 463-463.

Vollen L (2001) All that remains: identifying the victims of the Srebrenica Massacre. *Cambridge Quarterly of Healthcare Ethics* **10**, 336-340.

Wainscoat JS, Hill AV, Boyce AL, Flint J, Hernández M, Thein SL, Old JM, Lynch J.R., Falusi AG, Weatherall DJ & Clegg JB (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. *Nature* **319**, 491-493.

Wainwright FT (1962) *Archaeology and place-names and history: an essay on problems of co-ordination*. London: Routledge & Kegan Paul.

Waite ER, Child AM, Craig OE, Collins MJ, Gelsthorpe K & Brown TA (1997) A preliminary investigation of DNA stability in bone during artificial diagenesis. **168**, 547-554.

Wakeley J (1993) Substitution Rate Variation among Sites in Hypervariable Region-1 of Human Mitochondrial-DNA. *Journal of Molecular Evolution* **37**, 613-623.

Wall JD (2000) Detecting ancient admixture in humans using sequence polymorphism data. *Genetics* **154**, 1271-1279.

Walsh PS, Metzger DA & Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic samples. *BioTechniques* **10**, 503-516.

Wang L, Oota H, Saitou N, Jin F, Matsushita T & Ueda S (2000) Genetic structure of a 2,500-year-old human population in China and its spatiotemporal changes. *Molecular Biology and Evolution* **17**, 1396-1400.

Ward R, Frazier B, Dew-Jager K & Pääbo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. *Proceedings of the National Academy of Science USA* **88**, 8720-8724.

Ward R & Stringer C (1997) A molecular handle on the Neanderthals. *Nature* **388**, 225-226.

Watson E, Forster P, Richards M & Bandelt HJ (1997) Mitochondrial footprints of human expansions in Africa. *American Journal of Human Genetics* **61**, 691-704.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256-276.

Weale ME, Weiss DA, Jager RF, Bradman N & Thomas MG (2002) Y chromosome evidence for Anglo-Saxon mass migration. *Molecular Biology and Evolution* **19**, 1008-1021.

Weniger GC (1989) The Magdalenian in Western Central-Europe, settlement pattern and regionality. *Journal of World Prehistory* **3**, 323-372.

Whittam TS, Clark AG, Stoneking M, Cann RL & Wilson AC (1986) Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. *Proceedings of the National Academy of Science USA* **83**, 9611-9615.

Whittle A (1996) *Europe in the Neolithic*. Cambridge.

Wickam-Jones CR (1994) *Scotland's first settlers*. London: B.T.Batsford Ltd.

Wilkinson-Herbots HM, Richards MB, Forster P & Sykes BC (1996) Site 73 in hypervariable region II of the human mitochondrial genome and the origin of European populations. *Annals of Human Genetics* **60**, 499-508.

Williams P & Newman R (forthcoming) Excavations at Grove Farm, Market Lavington, Wiltshire, 1986-1990. The development of a Roman, Saxon and Medieval settlement. In *Wessex Archaeology Monograph*: Wessex Archaeology.

Wilson AC, Cann RL, Carr SM, George M, Gyllensten U, Helm-Bychowski KM, Palumbi SR, Prager EM, Sage RD & Stoneking M (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society* **26**, 375-400.

Wilson DM & Barsky D (2001) The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA. *Mutation Research-DNA Repair* **485**, 283-307.

Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N & Goldstein DB (2001) Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5078-5083.

Winick C (1970) *Dictionary of Anthropology*. London: Greenwood Press.

Wolpoff MH, Wu XZ & Thorne AG (1984) Modern Homo sapiens origins: A general theory of human evolution involving the fossil evidence from East Asia. In *The origins of modern humans: a world survey of the fossil evidence.*, pp. 411-483 [FH Smith and F Spencer, editors]. New York: Liss.

Wood J, Ferrell R & DeWitte-Avina S (2002a) The role of infectious disease in human populations: New findings from the Black Death. *American Journal of Physical Anthropology*, 167-168.

Wood JW, Ferrell RJ, DeWitte-Avina SN & Holman DJ (2002b) Mortality differentials during the Black Death in England, 1348-50: A reassessment. *American Journal of Human Biology* **14**, 90.

Woodward S, King M, Chiu N, Kuchar M & Griggs C (1994) Amplification of ancient nuclear DNA from teeth and soft tissues. *PCR methods and applications* **3**, 244-147.

Yang DY, Eng B, Waye JS, Dudar JC & Saunders SR (1998) Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Human Anthropology* **105**, 539-543.

Yorke B (1990) *Kings and Kingdoms of early Anglo saxon England*. London and New York.

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV & TylerSmith C (1997) Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. *American Journal of Human Genetics* **60**, 1174-1183.

Ziegler P (1969) *The Black Death*. London: Collins.

Zierdt H, Hummel S & Herrmann B (1996) Amplification of human short tandem repeats from medieval teeth and bone samples. *Human Biology* **68**, 185-199.

Zink A, Haas CJ, Reischl U, Szeimies U & Nerlich AG (2001) Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *Journal of Medical Microbiology* **50**, 355-366.

Zink AR, Reischl U, Wolf H & Nerlich AG (2002) Molecular analysis of ancient microbial infections. *Fems Microbiology Letters* **213**, 141-147.

Zouros E, Freeman KR, Ball AO & Pogson GH (1992) Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus*. *Nature* **359**, 412-414.

Zvelebil M (1986) *Hunters in transition*. Cambridge: Cambridge University Press.

Zvelebil M (1989) On the transition to farming in Europe, or what was the spreading with the Neolithic: a reply to Ammerman (1989). *Antiquity* **LXIII**, 379-383.

Zvelebil M & Zvelebil KV (1988) Agricultural transition and Indo-European dispersals. *Antiquity* **LXII**, 574-583.