

**Tracing the phylogeography of human populations in Britain based on 4<sup>th</sup>-11<sup>th</sup> century mtDNA genotypes.**

*A. Töpfung<sup>§</sup> M.T.P. Gilbert, <sup>‡¶</sup> J.P. Dumbacher,<sup>†</sup> and A.R. Hoelzel <sup>§\*</sup>*

<sup>§</sup> School of Biological and Biomedical Sciences, University of Durham, South Road, Durham, DH1 3LE, UK.

<sup>‡</sup> Ancient Biomolecules Centre, Departments of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

<sup>¶</sup> *Current address Ecology and Evolutionary Biology, University of Arizona, 1041 E. Lowell St., Tucson AZ 85271, USA.*

<sup>†</sup> Smithsonian Institution, NMNH - Genetics Program, 3000 Connecticut Ave., NW Washington, DC 20008-2537, USA.

\* correspondence; a.r.hoelzel@dur.ac.uk □ □

## **Abstract**

Some of the transitional periods of Britain during the first millennium AD are traditionally associated with the movement of people from continental Europe, composed largely of invading armies (e.g. the Roman, Saxon and Viking invasions). However, the extent to which these were migrations (as opposed to cultural exchange) remains controversial. We investigated the history of migration by women by amplifying mtDNA from ancient Britons who lived between approximately 300-1,000 AD, and compared these with 3,549 modern mtDNA database genotypes from England, Europe and the Middle East. The objective was to assess the dynamics of the historical population composition by comparing genotypes in a temporal context. Towards this objective we test and calibrate the use of rho-statistics to identify relationships between founder and source populations. We find evidence for shared ancestry between the earliest sites (predating Viking invasions) with modern populations across the north of Europe from Norway to Estonia, possibly reflecting common ancestors dating back to the last glacial epoch. This is in contrast with a late Saxon site in Norwich, where the genetic signature is consistent with more recent immigrations from the south, possibly as part of the Saxon invasions.

## **Introduction**

Our understanding of the British first millennium has come mainly from the study of historical, archaeological and linguistic data, e.g. written sources, grave-good contents and place-names. Based on this, several ideas have been proposed for the settlement of Britain. At one extreme, the population of Britain was founded by a series of migrations from Continental Europe, starting in the Holocene, and at the other, these transitional processes involved only trade and cultural exchange, rather than the movement of people (Dennell 1983, Dyer 1990, Hamerow 1997, Simon and Rigby 1997). The reality may have been somewhere in between, and this may be different for each of the historical transitions.

Inferences from modern genetic analyses have facilitated our understanding of the biological composition of the British population (Falsetti and Sokal 1993, Mastana and Sokol 1998, Wilson et al. 2001, Weale et al. 2002, Capelli et al. 2003). For example, analyses of Y-chromosomes suggest the influence in modern Britain of patriline genotypes from historical invading armies (especially from Anglo Saxon (Weale et al. 2002) and Viking (Wilson et al. 2001, Capelli et al. 2003) invasions). Yet, the role of women and their genetic footprint during these migrations is less well understood, largely due to the relatively poor resolution offered by human mtDNA for population assignment in Europe.

The modern distribution of genetic markers in Europe has been interpreted as reflecting waves of expansion over the last ~40,000 years, since the appearance of anatomically modern humans in Europe in the Late Upper Paleolithic. During the last glacial maximum (~18,000 YBP), northern populations are thought to have migrated south in search of better climate and shelter (Jochim 1983, Soffer 1987, Strauss 1990) and expanded from these refuges to re-populate the north when the climate improved. According to the archaeological record, Britain

was depopulated from 22,000 until about 13,000 BP, when the earliest human remains are again found. However, the time and exact route of this re-occupation is yet not well established, and may have been a punctuated rather than a continuous process (Housley et al. 1997). After re-settling, it is thought that Mesolithic people must have been constantly on the move in order to survive (Clark 1936, Fischer 1991). Later on, in the late Neolithic, their hunter-gatherer lifestyle was replaced by a more sedentary agricultural way of life. Farming had originated in the Near East around 10,000 YBP and expanded northwards into Europe, with some authors proposing that this spread was accompanied by a migration of people (Ammerman and Cavalli-Sforza 1971, Cavalli-Sforza et al. 1993).

In Great Britain, Y-chromosome data (Wilson et al. 2001, Weale et al. 2002, Capelli et al. 2003), and classical genetic markers (Falsetti and Sokal 1993, Mastana and Sokol 1998) showed evidence of genetic clines, suggesting a major immigration centred on the southeast. However, it is unclear whether this represents Saxon immigration or earlier events.

In this study we used available archaeological material and modern sequences to assess the genetic signature of England during the first millennia AD, and address the question of historical matrilineal origins as indicated by haplotype distribution and haplogroup frequencies during the Roman and Saxon periods. It was expected that a comparison of different time points could help identify patterns of population movement over time, based on a detailed analysis of haplotype frequencies. We were especially interested in a comparison of the early and late Saxon periods to investigate this transitional period.

## **Materials & Methods**

### **Samples**

Access to the human remains exhumed from five archaeological sites was gained for this study. These were Newarke Street at Leicester (Cooper 1996); Buckland at Dover, Norton at Cleveland Market (Sherlock and Welch 1992), Lavington at Salisbury and Castle Mall at Norwich. The excavations were all recent (between 1983 and 1993), and all samples have been kept in room temperature storage with minimal handling (only as needed for labelling and record keeping). Some teeth were from intact mandibles, and are not known to have been handled at all. This is important as studies have found a high proportion of contaminants on museum material (Malmstrom et al. 2005), but the contamination level varies (e.g. Noonan et al. 2005) presumably as a consequence of the extent of handling.

The site at Leicester was known to have been a Roman-British cemetery. At least 21 graves contained coffins (detected from the presence of nails), and their position was East to West. Although there were no grave goods, other deposited objects (e.g. coins) indicated a date of 4<sup>th</sup> century. The cemetery at Norton seems from its size and grave goods to have been associated with a small Anglo-Saxon community within the former Bernician kingdom. The date of this site was estimated to be 5<sup>th</sup> to 6<sup>th</sup> century based on the archaeological finds. A wide range and variety of objects were found at Buckland, consistent with the cosmopolitan nature of the Kentish kingdom during the sixth century. It apparently represents the traditional place of a wealthy and well-established local community. The archaeological evidence at Lavington suggests that the cemetery near this market site is from the early Saxon period (mid 6<sup>th</sup> to 7<sup>th</sup> century). 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 Late

Saxon cemetery was found beneath the southern bailey rampart. Ceramics from grave fills suggest that this cemetery was established in the late 10<sup>th</sup> century, with burial continuing into the mid 11<sup>th</sup> century. However, radiocarbon dates indicated materials from as early as the late 9<sup>th</sup> century (cal AD 890-1020). A total of 319 ancient dental samples were analysed from 156 individuals, and from these 721 DNA extracts were obtained for PCR amplification.

The comparative data set from modern populations of Europe for the founder and genetic distance analyses consisted of mtDNA HVS-I sequences from the following populations: Armenia (N=191), England (N=258), Estonia (N=149), northern France (N=101), Finland (N=176), Iceland (N=467), Norway (N=565), northern Germany (N=107), Palestine (N=117), Saami (N=176), Scotland (N=981), Spain (N=181) and Western Isles (N=181). Modern sequences were obtained from public databases, requested from the authors or deduced from the respective publications. A listing of the accession numbers and source references is available on request from the corresponding author.

#### Ancient DNA authentication

To ensure that the ancient DNA (aDNA) sequences obtained were authentic, we followed criteria recommended by Cooper and Poinar (2000). However, for the confirmation that DNA was present in the sample, we relied on replicate amplification both within and among labs as done in various other studies (e.g. Lalueza-Fox et al. 2004; Ricaut et al. 2005; Töpf & Hoelzel, 2005), rather than amino acid racemization or the amplification from a second species (none were available).

All ancient DNA extractions were carried out in a laboratory physically separated from the main building and exclusively dedicated to ancient DNA manipulation. PCR and post-PCR

analyses were carried out in the main laboratory. In addition, a one-way (from the ancient lab to the PCR lab) procedure was always followed to avoid the imperceptible carrying of DNA aerosols on clothes or skin into the ancient DNA laboratory (MacHugh et al. 2000).

To detect possible contamination, extraction, PCR and carrier negative controls were undertaken every 10 samples. To trace observed contamination, DNA sequences from the authors and other lab members were recorded for comparisons. Only independent extractions and amplifications (from different samples from the same skeleton) yielding identical sequences were included in the population analyses, and up to nine extracts from a given individual were amplified. Lesions in the aDNA template would be expected to be non-reproducible from different extracts, and artefact lesions at a given site (Gilbert et al. 2003), detectable across individuals or clones. Cloning was carried out for two samples, and 21 and 5 clones sequenced, respectively. Quantification of aDNA present was estimated by the number of cycles needed to obtain amplification (based on Rameckers et al. 1997). Four samples were replicated by three independent laboratories (one at the Ancient Biomolecules Centre, Oxford University, one at the University of Arizona and two at the Smithsonian Institution).

#### DNA extraction and PCR amplification

Teeth used for extraction were soaked with constant agitation in 30% bleach, rinsed in sterile water and 70% ethanol (30 min each). Samples were then UV-irradiated after drying (254 nm for 20 minutes on each face). They were wrapped in sterile metallic paper and broken open by a vice, so only internal parts of the tooth were used for DNA extraction. Dental pulp or its remnants adhered to the wall of the pulp chamber were collected by drilling using sterile equipment and collected into double-autoclaved tubes. DNA was extracted by means of a

sensitive method modified from Schmerer et al. (1999). Powdered dental material was incubated in lysis buffer (0.45M EDTA pH8, 0.5% SDS, 10mM Tris-HCl pH8) and  $0.65 \text{ mg ml}^{-1}$  Proteinase K final concentration for 24 h at 55 °C with constant agitation. After incubation, an aliquot was extracted twice with phenol/ chloroform/ isoamilalcohol. The DNA was then concentrated by adding 10  $\mu\text{l}$  of a silica suspension, which in the presence of an appropriate chaotropic agent (both from QIAex II gel extraction kit, Qiagen Ltd.) binds to the DNA, and can be retained and washed on a filtered spin column. DNA was eluted in alkaline buffer and stored at 4°C prior to PCR amplification to reduce the effect of inhibitory compounds (Montiel et al. 1997). A fragment of 264 bp (including primers) of the mtDNA HVS-I was amplified for 45 cycles using primers 16099 (5' AACCGCTATGTATTTTCGTAC3') and 16331 (5' TTTGACTGTAATGTGCTATGT A3') (numbering according to Anderson et al. 1981). PCR products were purified and DNA sequenced using the automated ABI Prism Dye Terminator system and run in an Applied Biosystems 373 DNA sequencer providing 207 bp readable sequence (the beginning of some sequences was unclear and therefore discarded) starting at position 16123.

#### Phylogenetic analysis

Sequences were aligned using the Sequencher 3.0 software (Gene Codes Corporation, USA) and polymorphic positions identified using MEGA 2.1. Reduced median networks (Bandelt et al. 1995) were constructed using NETWORK 2.0. To maximise the resolution, all of the segregating sites found along the 207 bp (including deletions) were used for the analyses.

#### Founder analysis

Genetic distances (first described on Forster et al. 1996) to putative source populations were calculated following Helgason and co-workers (Helgason et al. 2001). Rho estimates the mutational divergence between populations involved in founder events. It assumes that after a founder event the haplotypes observed in the new population will be a random sub-sample of haplotypes from the source population. The index is calculated as the average number of mutational differences between the sequences of the founder population and the closest sequences observed in the putative source population.

However, as rho distances were found to be dependent on the sample size of the source populations, a correction for sample size was applied (see Results). This consisted of a random sampling of the source populations so that sample sizes matched (with N=150). Choosing a cut-off of N=150 for the putative source populations restricted the pool of possible source populations to 12 (based on the database sequences available), but this was deemed to include enough of the key populations for the required analyses. Random numbers were generated using the 'Research Randomizer' at <http://www.randomizer.org>. A test of the impact of founder population sample size on rho was undertaken by randomly sub-sampling the modern England sample as a putative founder population. It was first sub-sampled for N=150, and this was compared against putative source populations with samples of N=150. The sample for England was then randomly sub-sampled for N=20, and this was repeated 20 times for each comparison against putative source populations (with population sample sizes of N=150). The mean and standard deviation for the 20 repeats of N=20 were calculated and compared against the data for the larger sample size (N=150).

Genetic distances and haplogroup assignment

It would be expected that for the relevant time period, most of the genetic change in the mtDNA pool would have been due the redistribution of lineages by migration and drift. Further, the shallow pattern of diversity for human mtDNA means that genetic distances based on nucleotide diversity are confounded by noise and reduced resolution. Therefore, genetic distances were calculated based on haplogroup frequencies.

Haplotypes were assigned to haplogroups (hg) and sub-haplogroups according to the West Eurasian mtDNA genealogy (based on Macaulay et al. 1999) using the following algorithm (numbering according to Anderson et al. 1981, minus 16000 for clarity): 126C assigned to hg J (and additional segregating site 261T assigned to sub-hg J1; 145A 231C 261T to sub-hg J1a; 145A 222T 261T to sub-hg J1b; 145A 172C 222T 261T to sub-hg J1b1 and 193T to sub-hg J2); 126C 294T assigned to hg T (and additional segregating sites 163G 186T 189C assigned to sub-hg T1; 304C to sub-hg T2; 292T to sub-hg T3; 324C to sub-hg T4 and 153A to sub-hg T5); 224C 311C assigned to hg K; 249C and either 189C or 327T assigned to hg U1; 129C assigned to hg U2; 270T assigned to hg U5 (and additional segregating sites 192T assigned to sub-hg U5a; 192T 256T to sub-hg U5a1; 256T to sub-hg U5a1a; 189C to sub-hg U5b and 189C 144C to sub-hg U5b1); 172C 219G assigned to hg U6 (and additional segregating site 278T to U6a and 311C to U6b); 318T assigned to hg U7; 298C assigned to hg V; 129A 223T assigned to hg I; 223T 292T assigned to hg W; 189C 223T 278T assigned to hg X and 223T assigned to 'IWX'. Sequences not matching any of the above were group together as 'other', though the polyphyletic nature of this grouping is likely to reduce resolution.

Genetic distances between populations based on these haplogroup frequencies were calculated using the *f* distance, based on the chord distance introduced by Cavalli-Sforza and Edwards (1967). The resulting matrix of geometric genetic distances was reduced to a 2-

dimensional space by means of a multidimensional scaling (MDS) analysis, using the SPSS 11.0 software.

## **Results**

### DNA extraction and authentication

A total of 48 DNA sequences could be authenticated and were included in the population analyses (accession numbers xxxx-xxxx). These included 6 (out of 15) from Leicester (3 females, 2 males and 1 unsexed adult), 12 (out of 23) from Norton (11 males and 1 unsexed child), 7 (out of 31) from Buckland (2 females and 5 males), 6 (out of 28) from Lavington (3 females and 3 unsexed children) and 17 (out of 59) from Norwich (7 females, 1 male and 9 unsexed children). Some of the variation in success rate among sites could be due to differential preservation. However, the number of extracts per sample was not the same for each site (depending on the size of tooth available), and the range of success was more even when this is taken into account (19% to 26%). DNA sequences were considered authentic when at least 2 independent extracts from different dental samples from the same individual yielded identical DNA sequences and all controls were negative. Matching sequence from separate teeth meant that the sequence was very unlikely to be the result of post-mortem changes to the DNA (and any consistent bias in the amplification (see Gilbert et al. 2003) should have been detected among individuals or clones). Although a level of contamination was detected in some negative controls and extracts, this could be traced to its origin, which was mostly from lab members (working in the ancient DNA lab).

Authenticated ancient samples showed appropriate aDNA molecular behaviour (such as signs of depurination and deamination lesions, arising due to oxidation and hydrolysis)

including chimeric sequences with a preponderance of C/T heteroplasmic positions (Hofreiter et al. 2001, Gilbert et al. 2003). This is observed as double or “heterozygous” peaks in the trace sequence and as differences among clones. Ancient template is also indicated by their failure to amplify an alternative 400bp PCR fragment (using conserved human primers for the mtDNA control region, Vigilant et al. 1989). No sequences showed obvious conflict with haplogroup-defining segregating sites, and all were consistent with modern European lineages. Cloned sequences also behaved as expected in aDNA work. In one case where the sequence was cloned in two overlapping fragments, all 5 clones showed two polymorphic sites while 3 clones showed a third site. These clones also showed some unique non-reproducible sites, presumably due to lesions in the DNA. In another case, two out of 21 clones had all 5 polymorphic sites that had been replicated from two other tooth extracts, while 19 were lab member sequences, or derivations of the expected sequence. Four samples were replicated among independent labs. The full sequences of two haplotypes were confirmed at Oxford and Arizona, and a 150bp subsequence (representing all polymorphic sites but one) was confirmed for two haplotypes at the Smithsonian (all four of these haplotypes were unique to the ancient sample set).

### Phylogenetic analyses

Reduced median networks (RMN) for the ancient populations are shown in Figure 1. Despite the short sequence analysed, most of the major Eurasian haplogroups and sub-haplogroups can be identified in both networks. However, a marked difference between the early and late Saxon sites can also be seen. The early site shows a large proportion of sub-

haplogroups U5a1 and U5a1a, and also I, V and W, while the later site at Norwich shows a large proportion of sub-haplogroup T1. Sub-haplogroup T1 is still relatively common in modern England, but absent from the 4<sup>th</sup> to 7<sup>th</sup> century sites (which present sub-haplogroup T4 and T5). The RMN for the early site also seemed more reticulated, with interconnected branches. The relationship between ancient and modern haplogroup frequencies in England is illustrated in Table 2. Although all the major haplogroups are represented in both populations, their proportion differ (Spearman's rank correlation:  $Z= 1.19$ ,  $p= 0.23$ ),

Only seven haplotypes were shared between two or more individuals, both at the same and different archaeological sites. Six of these haplotypes are frequent in modern populations (ranging from 1-18% of the full database), and therefore might be shared between non-related individuals. Just one haplotype was rare enough to possibly indicate maternal relationship – a sequence unique to Buckland shared between two adult male individuals.

#### Founder analysis

Rho distances (first described by Forster et al. 1996) were computed to quantify mutational differences between the ancient and modern populations, and by inference, identify possible source populations. For this, the method proposed by Helgason and co-workers (2001) was followed, where the authors analysed mtDNA from the islands of the North Atlantic to identify their proportion of Celtic and Viking ancestry. 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, rho distances will be affected by sample size. To investigate the extent of this sample size effect, rho distances were computed between the ancient population of England (used as the founder population) and random sub-

samples from our largest putative source population sample (Scotland, total N= 981; Figure 2). These subsamples varied in size from 50 to 800 sequences, and although the source population was always the same, rho values diminished with increasing sub-sample size. On the basis of this test we determined that it would be inappropriate to simply use the full sample sizes for all putative source populations, as these varied over nearly three orders of magnitude, and that a small standard sample of N=50 would also be inappropriate due to sampling effects. However, a sub-sample size of 150 showed a similar rho value to larger sub-sample sizes (due to saturation, increasing sample size resulted in smaller changes to rho beyond this point), and allowed the inclusion of 13 populations in our comparative analysis (see Table 1). Therefore, source population sub-samples were standardised at N=150 (with the exception of Estonia (N=148), Northern Germany (N=107), Northern France (N=101) and Palestine (N=117)), where the full sample was used.

The ancient sample was considered both as a whole, and subdivided according to time period. For the ancient population as a whole, the pattern of rho distances was similar to that seen for modern England (Pearson's rank correlation:  $Z = 2.56$ ,  $p = 0.024$ ), though there were apparent differences (Table 1). However, the late Saxon site in Norwich (9<sup>th</sup> -11<sup>th</sup> century) showed a distinct set of putative source populations compared to the earlier sites (three Saxon sites dated c. 5<sup>th</sup> -7<sup>th</sup> century and a Romano-British site from c. 4<sup>th</sup> century) (Table 1). All ancient British sample groups showed genetic similarity to the putative source sample from England, as expected. However, the four other putative source populations with the lowest rho values (greatest similarity) were all from north of England for the early ancient samples, and all from the south for the late Saxon sample from Norwich (Table 1, columns 2 and 3).

However, the sample sizes for the putative founder populations are small for these tests (N=31 for the 'early' and N=17 for the 'late' site). To test the magnitude of sampling error due to small sample size, we re-sampled the England sample (treating it as a putative founder population) for 20 individuals, 20 times, and compared the mean and s.d. with estimates based on a sample size of 150 (Table 1, columns 4 and 5). This was done simply to assess the likely extent of the error, and not to test the difference for statistical significance. We found good correspondence for the ordering of these estimates (Table 1), and at the extreme, sampling error would have had to shift rho values by 2-3 s.d. to generate by chance the pattern seen comparing the early and late ancient samples. We therefore conclude that the observed inversion of northern and southern putative source populations is not due to sampling effects.

The average magnitude of rho values was significantly greater for the ancient than the modern England populations (Mann-Whitney U-test,  $z = -2.725$ ,  $p = 0.006$ ), possibly reflecting greater recent integration (through both immigration and emigration) with the 13 putative source populations. The putative source populations with the lowest rho-values suggest the contribution of both northern and southern populations to the composition of modern England.

## Genetic distances

Genetic distances based on haplo- and sub-haplogroup frequencies were calculated and the resulting matrix reduced to two-dimensions by means of a MDS plot (Figure 3). This result is very consistent with our results for the rho distances. The ancient samples are well separated from the cluster of modern populations, however the late Saxon site shows closer genetic

distances to populations situated to the north (Scotland, Estonia, Norway, Finland and Western Isles) while the early Saxon site is closer to more southern populations (Germany and Spain, as well as the near eastern populations of Armenia and Palestine). The MDS data are also consistent with the reduced median networks shown in Figure 2, as the early and late ancient samples show differing representations of modern haplogroups in the network analysis (also seen in Table 2).

## **Discussion**

We compared ancient mtDNA sequences from two different periods and different locations in southern and central England with modern samples from Europe and the Middle East. We found a different representation of haplogroups in the two ancient samples grouped by period. Sub-haplogroup U5a1, part of haplogroup U5, which is believed to have originated or spread in Europe during the Upper Paleolithic, is well represented in the early ancient sample. On the other hand, the late ancient site shows a large proportion of sub-haplogroup T1, which is associated with the Neolithic expansion (Richards et al. 2002). This suggests a different composition of the two temporal samples, which may be associated with different geographic origins. Although our sample sizes did not permit a test of geographic differentiation for a given time period, the geographic distances are small, and the time frame reflects a period when there are known to have been major demographic changes. We therefore emphasize the comparison between early and late Saxon sites.

The sample from northern Germany, representing the geographic region from which some Saxon groups originated, was not among the genetically closest founder populations for the early Saxon group (based on the rho data). This is in agreement with modern Y-chromosome

data that indicated a limited continental input from north Germany in southeast England, where most of the Anglo-Saxon settlements were located (Capelli et al. 2003). Our founder analyses showed a clear inversion of the genetically closer putative source populations for the early and late Saxon British sites. Northern populations were closest to the early ancient sites, while most southern populations were closer to the late Saxon site (Northern France showed unexpectedly high rho values, even against modern England). Although re-sampling analyses showed potential bias using the rho statistic both for varying source population size and for small putative founder population size, we demonstrate the magnitude of these biases and find a well-supported difference even after taking these factors into account. The difference was reinforced in the MDS analysis where the early and late samples did not cluster together, but clustered closer to northern and southern samples, respectively. The geographic relationship of these populations (as opposed to historical connection due to early medieval migration/invasion), seems to suggest deeper common ancestry among northern populations across Europe. This supports new findings as well as some new interpretations of earlier archaeological data suggesting a common Mesolithic culture from Britain to Estonia (Clark 1936, Coles 1998, Pettitt 2003).

According to some archaeological interpretations based on the scarcity of archaeological findings, northern Europe was abandoned as people moved to southern refuges during the peak of the last glacial maximum (Mellars 1974, Evans 1975), from where they re-expanded north when the climate improved (Housley et al. 1997). However, others have proposed that some people could have stayed in the north, especially in areas that are now submerged under the North Sea (Coles 1998). When the Scandinavian and British ice sheets reached their maximum extent, and the North Sea as a consequence receded to its lowest level (Fairbanks 1989), Britain

was connected to the continent by a land bridge. This dry land, referred to as Doggerland, is now believed to have lasted longer and been larger than previously thought, and may have been inhabited (Wymer 1991, Coles 1999). After the ice receded, they could have radiated out from those regions into communities further north (illustrated in Figure 4). Some of those now in the north of Europe (e.g. northern Germany) may represent people who migrated from the south, while some people in that region historically moved further north.

Coles (1999) proposed that the occupation of this land bridge may have played a role in delaying the onset of the Neolithic in Britain and Scandinavia, as a consequence of encounters between the Mesolithic northern populations and the agriculturist newcomers. Case (1969) in reference to the mixture of different traits of the British Neolithic, suggested that some of the donor cultures may have lived in coastal regions that are now submerged or eroded. Modern populations in the UK and Scandinavia may therefore share ancestry with people who lived in this region perhaps as recently as 7,500 YBP. This might also explain the presence of specific Neolithic haplotypes (such as J1b1) exclusive to the British Isles and Scandinavia (Helgason et al. 2001, Topf 2003), previously thought to represent posterior links between these populations.

Common ancestry could of course reflect common origins anytime between the LGM and Saxon periods, but we focus on the period when Doggerland was exposed due to the supporting archaeological evidence and the opportunity for northward expansion availed by the land bridge. McEvoy et al. (2005) also suggested that the genetic landscape of southeast Britain may have been shaped by older links with the continent, when Doggerland still existed. Genetic continuity has been proposed to exist among northern populations for other species as well (e.g. for *Rana lessonae*; Snell et al. 2005 ).

The different apparent source populations at the Norwich site may reflect a greater influence of postglacial migrations from the south and Near East in this region, or this particular site, or during this time period. This site was also likely under Danelaw at this time and the influence of the Danish Vikings. Modern England appears to continue to reflect these influences, though further Viking invasions and later immigration and emigration would have influenced the modern genotypic pattern as well. Modern data on Y-chromosome analysis has also suggested that the Danes had a great impact in this area (Capelli et al. 2003). Immigration of women to Norwich at this time is suggested by the fact that modern England retains hg T1 (at 2.5%), which is found in Norwich (23%) but absent from the early Saxon sites. Although the site at Norwich is likely to have been larger and more cosmopolitan than the other sites in our analysis, there is nothing about the cemetery at Castle Mall to indicate that it represented anyone other than people local to the Norwich area at that time.

Our analyses of mtDNA sequences from ancient and modern Europe show a distinct pattern for the different time periods sampled. Unfortunately the late ancient sample size is relatively small, but the resulting pattern is non-random, and seems to support archaeological findings. The prevailing theory for numerous species is that populations in Europe expanded from southern refugia after the last glacial maximum (see Hewitt 2000), though evidence exists for various northern refugia as well (see review in Stewart & Lister 2001). For humans, modern mitochondrial DNA data has indicated the importance of the Franco-Cantabrian refuge during the glacial period (Achilli et al. 2004, Pereira et al. 2005). However, this is not necessarily in conflict with the interpretation of our data. We present data that for the first time analyses genotypes from the pre-Saxon/ early Saxon period, and these are uniquely able to assess signatures for relatedness from this period of demographic change. A similar study using ancient

DNA provided new information on the source of founding populations in the Canary Islands (Maca-Meyer et al. 2004). Even if the majority of the human population of northern Europe emigrated south during the LGM, a proportion remaining behind could be responsible for the genetic signature we have identified.

#### Acknowledgements

We thank Rob Fleischer and Alan Cooper for assistance in replicating several of our ancient amplifications and Trevor Anderson, Mandy Marlow, Lorraine Mepham, John Lucas and Liz Popescu for providing access to the archaeological material. We thank Laurent Excoffier and Peter Forster for useful comments and discussions and Vincent Dubut for providing sequences from France.

#### References:

- Achilli, A., C. Rengo, C. Magri, V. Battaglia, A. Olivieri, R. Scozzari, F. Cruciani, M. Zeviani, E. Briem, V. Carelli, P. Moral, J. M. Dugoujon, U. Roostalu, E. L. Loogvali, T. Kivisild, H. J. Bandelt, M. Richards, R. Villems, A. S. Santachiara-Benerecetti, O. Semino, and A. Torroni. 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am. J. Hum. Genet.* **75**:910-918.
- Ammerman, A. J., and L. L. Cavalli-Sforza. 1971. Measuring the rate of spread of early farming in Europe. *Man* **6**:674-688.
- Anderson, S., A. Bankier, B. Barrell, M. Debruijn, A. Coulson, J. Drouin, I. Eperon, D. Nierlich, B. Roe, F. Sanger, P. Schreier, A. Smith, and I. Young. 1981. Sequence organisation of the human mitochondrial genome. *Nature* **290**:457-465.
- Bandelt, H. J., P. Forster, B. C. Sykes, and M. B. Richards. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* **141**:743-753.
- Capelli, C., N. Redhead, J. K. Abernethy, F. Gratrix, J. Wilson, T. Moen, T. Hervig, M. Richards, M. P. Stumpf, P. A. Underhill, P. Bradshaw, A. Shaha, M. Thomas, N. Bradman, and D. B. Goldstein. 2003. A Y-chromosome census of the British Isles. *Curr. Biol.* **13**:979-984.
- Case, H. 1969. Neolithic explanations. *Antiquity* **XLIII**:176-186.
- Cavalli-Sforza, F., and A. W. F. Edwards. 1967. Phylogenetic analyses: models and estimation procedures. *Am. J. Hum. Genet.* **19**:233-257.

- Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1993. Demic expansions and human evolution. *Science* **259**:639-346.
- Clark, J. G. D. 1936. *The Mesolithic settlement of northern Europe: a study of the food-gathering peoples of northern Europe during the early post-glacial period*. Cambridge University Press, Cambridge.
- Coles, B. J. 1998. Doggerland: a speculative survey. *Proc. Prehist. Soc.* **64**:45-81.
- . 1999. Doggerland's loss and the Neolithic. Pages 51-57 in B. J. Coles, J. Coles, and M. Schou Jørgensen, editors. *Bog bodies, sacred sites and wetland archaeology*. WARP occasional paper, Exeter.
- Cooper, A., and H. N. Poinar. 2000. Ancient DNA: Do it right or not at all. *Science* **289**:1139-1139.
- Cooper, L. 1996. A Roman cemetery in Newarke Street, Leicester. *Trans. Leicestershire Arch. Hist. Soc.* **LXX**:1-90.
- Dennell, R. 1983. *European economic prehistory: a new approach*. London.
- Dubut, V., Chollet, L., Murail, P., Cartault, F., Beraud-Colomb, E., Serre, M. and Mogentale-Profizi, N. 2004. mtDNA polymorphisms in five French groups: importance of regional sampling. *Eur. J. Hum. Genet.* **12**:293-300.
- Dyer, J. 1990. *Ancient Britain*. Batsford, London.
- Evans, B. 1975. *The environment of early man in the British Isles*. University of California Press, Berkeley.
- Fairbanks, R. G. 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, A. B., and R. R. Sokal. 1993. Genetic structure of human populations in the British Isles. *Ann. Hum. Biol.* **20**:215-229.
- Fischer, A. 1991. Pioneers in deglaciated landscapes: the expansion and adaptation of Late Paleolithic societies in Southern Scandinavia. Pages 100-122 in N. Barton, A. J. Roberts, and D. A. Roe, editors. *The Late Glacial in north-west Europe: Human adaptation and environmental change at the end of the Pleistocene*. CBA Research Reports, Oxford.
- Forster, P., R. Harding, A. Torroni, and H. J. Bandelt. 1996. Origin and evolution of native American mtDNA variation: A reappraisal. *Am. J. Hum. Genet.* **59**:935-945.
- Gilbert, M. T., A. J. Hansen, E. Willerslev, L. Rudbeck, I. Barnes, N. Lynnerup, and A. Cooper. 2003. Characterization of genetic miscoding lesions caused by postmortem damage. *Am. J. Hum. Genet.* **72**:48-61.
- Hamerow, H. 1997. Migration theory and the Anglo-Saxon "identity crisis". Pages 33-44 in J. Chapman and H. Hamerow, editors. *Migrations and invasions in archaeological explanation*. BAR International Series, Oxford.
- Helgason, A., E. Hickey, S. Goodacre, V. Bosnes, K. Stefansson, R. Ward, and B. Sykes. 2001. mtDNA and the islands of the North Atlantic: Estimating the proportions of Norse and Gaelic ancestry. *Am. J. Hum. Genet.* **68**:723-737.
- Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**:907-913.
- Hofreiter, M., V. Jaenicke, D. Serre, A. von Haeseler, and S. Pääbo. 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29**:4793-4799.
- Housley, R. A., C. S. Gamble, M. Street, and P. Pettitt. 1997. Radiocarbon evidence for the lateglacial human recolonisation of northern Europe. *Proc. Prehist. Soc.* **63**:25-54.

- Jochim, M. 1983. Paleolithic cave art in ecological perspective. Pages 212-219 in G. Bailey, editor. *Hunter-gatherer economy in prehistory*. Cambridge University Press, Cambridge.
- Laleuza-Fox, C., M.L. Sampietro, M.T.P. Gilbert, L. Castri, F. Facchini, D. Pettener and J. Bertranpetit. 2004. Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient Central Asians. *Proc. Royal Soc. B* **271**:941-947.
- Maca-Meyer, N., Arnay, M., Rando, J.C., Flores, C., Gonzalez, A.M., Cabrera, V.M. and Larruga, J.M. 2004. Ancient mtDNA analysis and the origin of the Guanches. *Eur. J. Hum. Genet.* **12**:155-162.
- Macaulay, V., M. Richards, E. Hickey, E. Vega, F. Cruciani, V. Guida, R. Scozzari, B. Bonne-Tamir, B. Sykes, and A. Torroni. 1999. The emerging tree of west Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **64**:232-249.
- MacHugh, D. E., C. J. Edwards, J. F. Bailey, D. R. Bancroft, and D. Bradley. 2000. The extraction and analyses of ancient DNA from bone and teeth: a survey of current methodologies. *Ancient Biomol.* **3**:81-102.
- Malmstrom, H., J. Stora, L. Dalen, G. Holmlund, and A. Gotherstrom. 2005. Extensive contamination in extracts from ancient dog bones and teeth. *Mol. Biol. Evol.* doi: 10.1093/molbev/msi195.
- Mastana, S. S., and R. J. Sokol. 1998. Genetic variation in the East Midlands. *Ann. Hum. Biol.* **25**:43-68.
- Mellars, P. 1974. The Paleolithic and Mesolithic. Pages 41-99 in C. Renfrew, editor. *British Prehistory*. Noyes Press, Park Ridge, New Jersey.
- Montiel, R., A. Malgosa, and E. Subirà. 1997. Overcoming PCR inhibitors in ancient DNA extracts from teeth. *Ancient Biomol.* **1**:221-225.
- Noonan, J.P., Hofreiter, M., Smith, D., Priest, J.R., Rohland, N., Rabeder, G., Krause, J., Dettler, J.C., Paabo, S. and Rubin, E.M. 2005. Genomic sequencing of Pleistocene cave bears. *Science*, **309**:597-600.
- Pereira, L., M. B. Richards, A. Goios, A. Alonso, C. Albarran, O. Garcia, D. M. Behar, M. Golge, J. Hatina, L. Al-Gazali, D. G. Bradley, V. Macaulay, and A. Amorim. 2005. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res.* **15**:19-24.
- Pettitt, P. 2003. Paleolithic cave art found in Britain. *Before Farming* **10**:1-2.
- Rameckers, J., S. Hummel, and B. Herrmann. 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.
- Ricaud F. X., A. Fedoseeva, C. Keyser-Tracqui, E. Crubezy, and B. Ludes. 2005. Ancient DNA analysis of human neolithic remains found in northeastern Siberia. *Am. J. Phys. Anthropol.* **126**:458-462.
- Richards, M., V. Macaulay, A. Torroni, and H. J. Bandelt. 2002. In search of geographical patterns in European mitochondrial DNA. *Am. J. Hum. Genet.* **71**:1171-1174
- Schmerer, W., S. Hummel, and B. Herrmann. 1999. Optimized DNA extraction to improve reproducibility of short tandem repeat genotyping with highly degraded DNA as target. *Electrophoresis* **20**:1712-1716.
- Sherlock, S. J., and M. G. Welch. 1992. *An Anglo-Saxon cemetery at Norton, Cleveland*. Council for British Archaeology, London.
- Simon, J., and V. Rigby. 1997. *Britain and the Celtic Iron Age*. British Museum, London.

- Snell, C., J. Tetteh, and I. H. Evans. 2005. Phylogeography of the pool frog (*Rana lessonae* Camerano) in Europe: evidence for native status in Great Britain and for an unusual postglacial colonization route. *Biol. J. Linnean Soc.* **85**:41-51
- Soffer, O. 1987. Upper Paleolithic Connubia, refugia and the archaeological record from Eastern Europe. Pages 333-348 *in* O. Soffer, editor. *The Pleistocene World - Regional Perspectives*. Plenum Press, New York.
- Steward J. R., and A. M. Lister. 2001. Cryptic northern refugia and the origins of the modern biota. *Trends Ecol. Evol.* **16**:608-613
- Strauss, L. G. 1990. The Last Glacial Maximum in Cantabrian Spain: the Salutrean. Pages 89-108 *in* O. Soffer and C. Gamble, editors. *The world at 18,000 BP. Volume 1: High latitudes*. Unwin Hyman Inc., London.
- Töpf, A.L. 2003. Mitochondrial DNA diversity and origin of human communities from 4<sup>th</sup>-11<sup>th</sup> century Britain. PhD Thesis, University of Durham.
- Töpf A. L., and A. R. Hoelzel. 2005. A Romani mitochondrial haplotype in England 500 years before their recorded arrival in Britain. *Biology Letters*. doi:10.1098/rsbl.2005.0314.
- Vigilant, L., R. Pennington, H. Harpending, T. D. Kocher, and A. Wilson. 1989. Mitochondrial DNA sequences in single hairs from southern African population. *Proc. Natl. Acad. Sci. USA* **86**:9350-9354.
- Weale, M. E., D. A. Weiss, R. F. Jager, N. Bradman, and M. G. Thomas. 2002. Y chromosome evidence for Anglo-Saxon mass migration. *Mol.Biol.Evol.* **19**:1008-1021.
- Wilson, J. F., D. A. Weiss, M. Richards, M. G. Thomas, N. Bradman, and D. B. Goldstein. 2001. Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc. Natl. Acad. Sci. USA* **98**:5078-5083.
- Wymer, J. 1991. *Mesolithic Britain*. Princes Risborough, Shire.

## Figure Legends:

Figure 1: a) Reduced median network for the ancient populations of Britain belonging to the early period, including both the early Saxon sites and the Romano-British settlement of Leicester (4th century). Samples belonging to the different settlements are represented as: (R) Leicester; (N) Norton; (L) Lavington and (B) Buckland. (b) Reduced median network for the late Saxon, the settlement of Norwich (9th-11th century). Node sizes are proportional to the number of individuals. Major haplogroups are indicated. Polymorphic sites defining haplo- and sub-haplogroups are underlined.

Figure 2: Effect of sample size of the source population in the computation of rho distances. For the same founder population (ancient England), rho distances were calculated to the same source population (Scotland), randomly re-sampled for different sample sizes.

Figure 3: Multidimensional scaling (MDS) plot of  $f$  distances (Cavalli Sforza and Edwards, 1967) based on haplogroup frequencies. The fifteen dimensions of the genetic distance matrix were reduced to two dimensions, accounting for the 96% of the variation defined by the original distance matrix.

Figure 4: Hypothetical extension of the land bridge, i.e. Doggerland, during the Holocene/late Mesolithic (~10,000 YBP) and hypothetical expansion routes (modified from Coles, 1999).

Table 1:

Rho distances between the ancient populations and several modern populations

Ancient (48)		Early ancient <sup>c</sup> (31)		Late ancient <sup>d</sup> (17)		England <sup>e</sup> (20)		England (150)	
Pop <sup>a</sup>	Rho <sup>b</sup>	Pop	Rho	Pop	Rho	Pop	Rho $\pm$ s.d.	Pop	Rho
EN	0.729	ES	0.581	PA	0.588	ES	0.595 $\pm$ 0.168	ES	0.587
ES	0.771	EN	0.613	EN	0.941	SC	0.615 $\pm$ 0.158	SC	0.596
PA	0.771	NO	0.731	NG	0.941	NG	0.671 $\pm$ 0.184	NG	0.647
NO	0.889	FI	0.742	AR	1.000	SP	0.686 $\pm$ 0.201	SP	0.660
FI	0.896	SC	0.812	SP	1.059	NO	0.719 $\pm$ 0.203	NO	0.704
NG	0.896	SP	0.839	ES	1.118	PA	0.740 $\pm$ 0.196	PA	0.713
SP	0.917	NG	0.871	SC	1.157	FI	0.740 $\pm$ 0.221	FI	0.720
SC	0.934	PA	0.871	FI	1.176	WI	0.817 $\pm$ 0.251	WI	0.807
AR	1.083	WI	0.935	NO	1.176	NF	0.838 $\pm$ 0.292	AR	0.807
WI	1.167	IC	0.989	WI	1.588	AR	0.843 $\pm$ 0.198	NF	0.827
IC	1.215	AR	1.129	IC	1.627	IC	0.925 $\pm$ 0.276	IC	0.904
NF	1.375	NF	1.129	NF	1.824	SA	1.460 $\pm$ 0.386	SA	1.453
SA	2.125	SA	1.677	SA	2.941				

<sup>a</sup>AR=Armenia, EN=England, ES=Estonia, FI=Finland, NF=Northern France, IC=Iceland, NG=North Germany, NO=Norway, PA=Palestine, SA=Saami, SC=Scotland, SP=Spain, WI=Western Islands. <sup>b</sup>Rho distances between the ancient populations (founder) and several modern populations (sources), with subsamples of 150. <sup>c</sup>4<sup>th</sup>-7<sup>th</sup> century <sup>d</sup>9<sup>th</sup>-11<sup>th</sup> century <sup>e</sup>The England putative founder results constitute a test of the method, showing the mean and s.d. for 20 sub-samples of 20 individuals, compared to a sub-sample of 150.

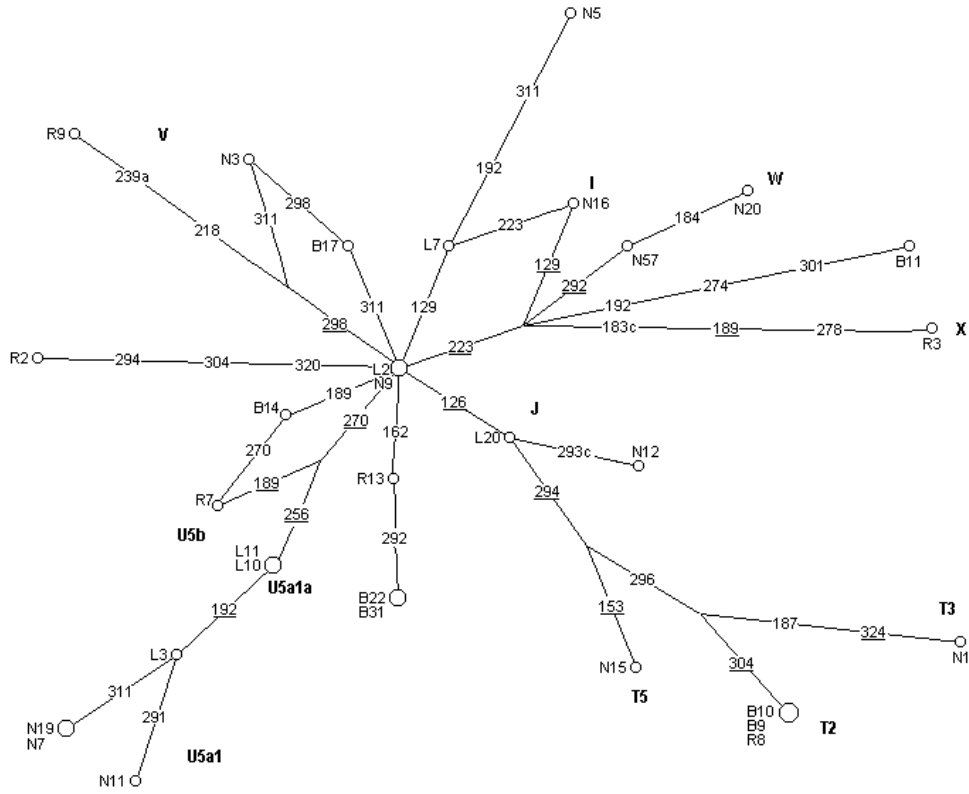
Table 2:

Haplogroup frequencies for the modern population of England and the ancient sample (also divided into early and late Saxon sites).

hg	Modern England	Ancient	Early	Late
J	0.078	0.083	0.065	0.118
J1	0.004	-	-	-
J1a	0.023	-	-	-
J1b	-	-	-	-
J1b1	0.012	-	-	-
J2	0.004	-	-	-
T	0.019	-	-	-
T1	0.023	0.083	-	0.235
T2	0.039	0.104	0.097	0.118
T3	0.004	-	-	-
T4	0.004	0.021	0.032	-
T5	0.008	0.021	0.032	-
K	0.066	0.021	-	0.059
U1	0.008	-	-	-
U2	0.012	-	-	-
U5	0.008	-	-	-
U5a	0.016	-	-	-
U5a1	0.027	0.083	0.097	-
U5a1a	0.004	0.042	0.097	-
U5b	0.012	0.042	0.032	0.059
U5b1	-	-	-	-
U6	-	-	-	-
U7	-	-	-	-
V	0.039	0.042	0.065	-
I	0.023	0.021	0.032	-
W	0.012	0.042	0.065	-
X	0.008	0.042	0.032	0.059
IXW	0.019	0.042	0.032	0.059
OTHER	0.531	0.313	0.323	0.294

Figure 1:

**a**



**b**

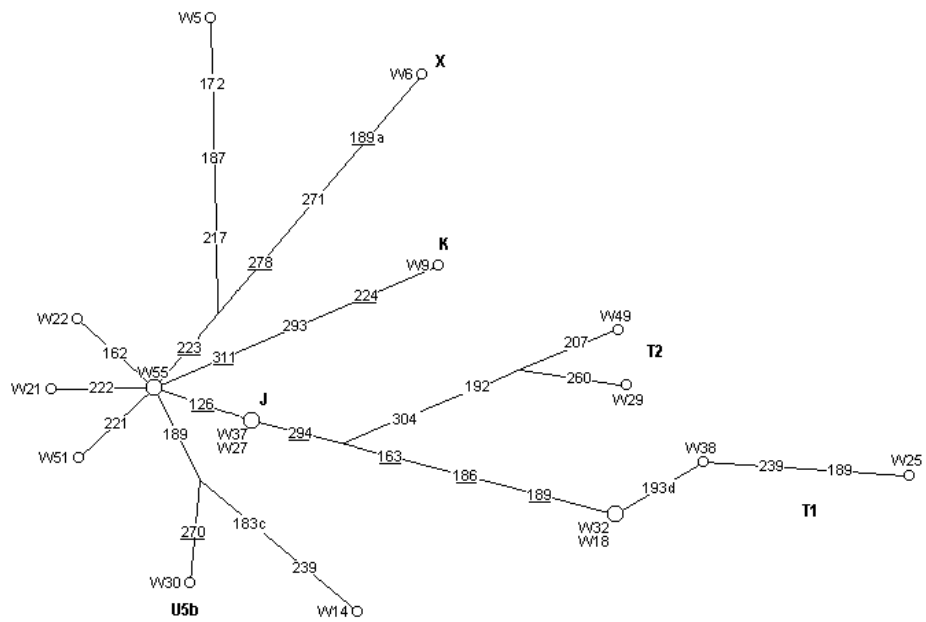


Figure 2:

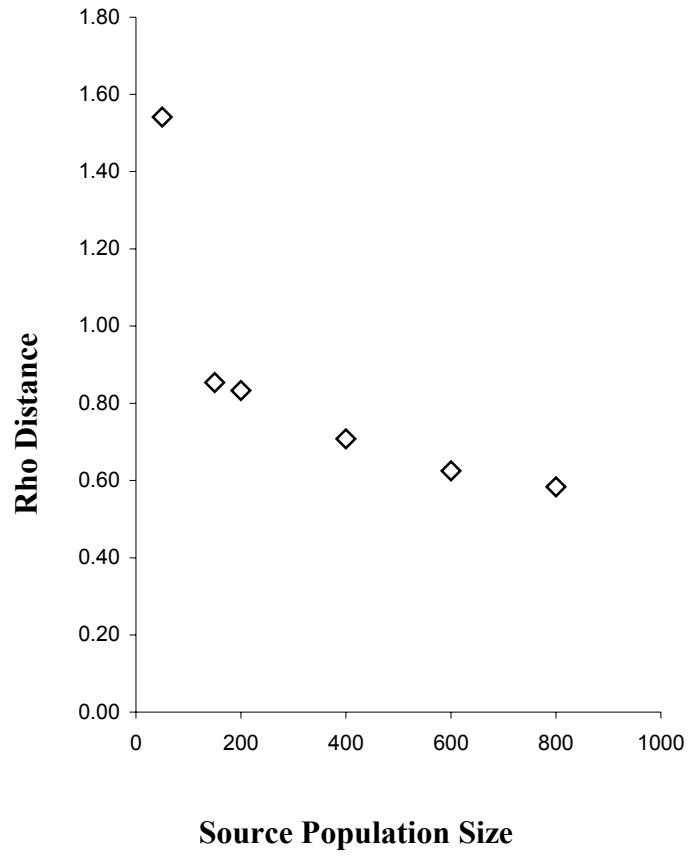


Figure 3:

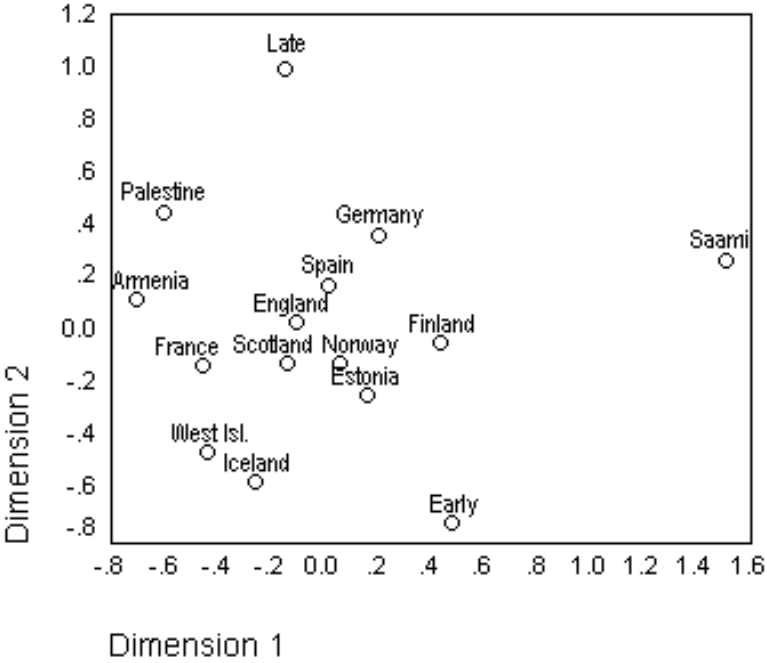


Figure 4:

