Using mtDNA to evaluate pioneer colonization scenarios for early prehistoric southern Scandinavia

Felix Riede*1, Marie Louise Stig Sørensen2 & Hans Eiberg3

*Corresponding author (f.riede@hum.au.dk).

DNA from living human populations can be used to infer their evolutionary and demographic histories, especially regarding initial dispersal events and subsequent population expansions. Southern Scandinavia was re-colonised by Late Palaeolithic hunter-gatherers beginning around 14 700 years BP, and the recent literature offers two competing hypotheses for the origins, direction and timing of this dispersal event. We present here the results of a study of maternally inherited mtDNA in 189 unrelated Danes, 64 of whom come from the island of Als and the remainder from various regions throughout the country. The mtRadius phylogeographic analysis tool is used to evaluate competing scenarios for the pioneer colonization of southern Scandinavia after the Last Ice Age. Our results are more consistent with an overall south-western dispersal trajectory for the maternally inherited mtDNA lineages. This contrasts with the current interpretation of the dispersal history of the paternally inherited Y-chromosome. The discrepancy can be reconciled if more complex demographic scenarios are taken into consideration.

Keywords: mtDNA, mtRadius, southern Scandinavia, human dispersal

Introduction

The last 25 years have witnessed a steep rise in the availability of human population genetic data for non-medical and non-forensic uses (Jobling et al. 2004; Pakendorf & Stoneking 2005). This has resulted in the emergence of a research field sometimes referred to as “archaeogenetics” (Renfrew & Boyle 2000), in which genetic data, most commonly focused on the selectively neutral mitochondrial inheritance system (mtDNA) or the non-recombining Y-chromosome (NRY), have been used to infer past demographic events such as initial population dispersals and expansions. One of the most thoroughly investigated of these events is the re-expansion of human hunter-gatherers into the high latitudes of Europe after the Last Ice Age (Forster 2004). Southern Scandinavia (northern Germany, Denmark and Scania), as a peninsular corridor to northern Europe, is an important geographical component in these hypotheses (Fig. 1). Insufficient published genetic data are available on this region, however, despite the fact that both geneticists (Hewitt 1999, 2000, 2001) and archaeologists (Price 1991:185) have noted that “northern Europe is an extraordinary laboratory for the investigation of human colonization and adaptation”.

Two distinct and opposing hypotheses presented in the following section regarding the geographical origin of the first human occupants of Scandinavia, the primary trajectory of dispersal and its timing can be found in the literature. We present here the results of a study of maternally inherited mtDNA in 189 unrelated Danes, 64 of whom come from the island of Als and the remainder from regions throughout the country. We describe the genetic variation in this sample and analyse it using the mtRadius phylogeographic analysis tool (Röhl et al. 2001). These data can be used to evaluate scenarios for the pioneer colonization of southern Scandinavia after the Last Ice Age. Our findings are more consistent with
dispersal from the south-west, but we also note that the dispersal trajectories of the maternally inherited mitochondrial DNA diverge somewhat from those of the paternally inherited Y-chromosome. We suggest that multiple dispersals and sex-specific bottlenecks at, and after, the initial colonization may best explain these patterns. It is recommended that future studies should focus on producing high-resolution NRY, whole-genome mtDNA and aDNA data in order to address these outstanding issues.

Two hypotheses for the pioneer colonization of southern Scandinavia

The literature on the earliest human re-colonization of Scandinavia offers two hypotheses, which will here be termed the eastern (H1) and western (H2) dispersal hypotheses, indicating the direction from which the populations are thought to have come. Each of these hypotheses has demographic (genetic) and cultural (archaeological) components and thus generates a number of predictions for the patterns which should be observed in the genetic and archaeological records:

H1: Eastern dispersal: Archaeological finds in southern Scandinavia dating back to the Late Palaeolithic became known through the excavations of Rust (1937, 1943), who also suggested a complicated re-colonization scenario involving several dispersal episodes. Based on similarities in the archaeological record between southern Scandinavia and the Ukraine, he postulated an eastern origin for the Hamburgian culture. While most archaeologists to date tend to favour a western dispersal model, a recent comprehensive review of the radiocarbon dating record for the earliest occupation of Scandinavia points to the east as a possible point of entry (Grimm & Weber 2008). The prediction generated by this model is that population links exist between eastern Europe (the Balkans and/or the Ukrainian refugium) and Scandinavia.

H2: Western dispersal: This hypothesis emerged first out of a growing body of archaeological evidence that linked the pioneer (Hamburgian) culture of southern Scandinavia with south-western Europe (see Bosinski 1982; Schmider 1982). More recently, a similar scenario has been offered by Otte (1990, 2000), although he suggests an initial entry from the west associated with the Hamburgian culture, followed by a second, later dispersal from the east associated with the Ahrensburgian culture. The identification of a significant pre-agricultural contribution to the modern European gene pool (Torroni et al. 1998; Richards et al. 2000; Torroni et al. 2001) eventually led to the emergence of a strong western dispersal hypothesis (Forster 2004; Gamble et al. 2004, 2005, 2006). The prediction of this model is a signature of western affinities of the southern Scandinavian population, linking it to the Franco-Cantabrian refugium.

In addition, it should be noted that some scholars (e.g. Töpf et al. 2006) would like to place the origin of the colonizing populations in the centre of northern Europe, the now-submerged North Sea land area, commonly referred to as Doggerland. This hypothesis rests on ancient DNA (aDNA) and modern population genetics. When examining aDNA from pre- and post-Anglo-Saxon contexts in Scandinavia and the British Isles, Töpf and colleagues attempted to rule out more recent historical mixing between these populations and argue for deep similarities with a shared central source population. The potential role of Doggerland in the Late Palaeolithic has long been acknowledged (see Coles 1998; Fuglestved 2005), although Bjerck (1995) has argued that this area may have been rather unsuited to human settlement and evidence of human activity from the now submerged area does not predate the Mesolithic (Andersen 2005; Glimmerveen et al. 2006; Mol et al. 2006; Gaffney et al. 2007). It therefore remains difficult to evaluate the hypothesis of Töpf and colleagues any further until definite archaeological evidence for the Late Glacial settlement of Doggerland is forthcoming. By the same token, however, it is noteworthy that evidence for the possible presence
of Hamburgian hunters in present-day Scotland has recently been presented (Ballin et al. 2010), indicating that an investigation of the early Late Glacial long-distance connections in an east-west direction, i.e. across Doggerland is a pressing issue for future research.

In the following we will discuss the mitochondrial genetic variation in the present Danish sample in light of our phylogeographic analysis, in the context of other genetic studies and in relation to the archaeological and climatic records.

Samples and methods
The sample is composed of 189 unrelated Danes of both sexes. 64 of the samples were collected by MLSS on the southern Danish island of Als as part of a project examining the long-term history of that region (Sørensen et al. 2001), while a further 21 samples were collected from Danish residents in Cambridge, England, by FR. Informed consent was obtained from all the participants prior to sampling. The remaining Danish samples were provided by HE and had been collected previously for medical research from volunteers then living in the Copenhagen area. In order to evaluate the extent of very recent population movements following the industrial revolution, each participant was asked to provide a maternal (and paternal) geographical and linguistic family history up to three generations ago wherever possible. Many studies of modern mtDNA do not differentiate between long-term residents and recent immigrants, as samples are obtained either from forensic contexts or from military recruits, for instance. Knowledge of such family histories not only allows the screening of a given set of samples, but it also facilitates a more detailed phylogeographic analysis. The overwhelming majority of participants were able to provide information on at least their parents’ area of origin, and many could trace their origin three generations back. Only those individuals who, to the best of their knowledge, had family roots in southern Scandinavia were asked to provide samples.

All the samples with the exception of those from the Panum Institute, which were provided pre-extracted, were obtained using a buccal swab, and DNA was extracted after air-drying overnight. The Chelex method (see Walsh et al. 1991) was used for DNA extraction, as it has proved to be the most efficient and most economical means of DNA extraction for PCR analysis from a variety of samples, including buccal swabs (Suenaga & Nakamura 2005). The collection of human DNA samples from cheek swabs is non-invasive, fast and effective, and is thus the method of choice for studies of this kind (Quinque et al. 2006). The following protocol was applied throughout: Part of the cotton end of the swab was cut off using sterilized scissors and placed into a 2 ml screw-top tube. 1 ml of distilled H$_2$O was added and, after vortexing, the samples were incubated for 15–30 min at room temperature. They were then centrifuged at 6000 rpm for five minutes and the supernatant removed, leaving about 20–30 μl in the tube. A volume of 170 μl of Chelex® slurry at a 5% concentration and 50 μl of Proteinase K (2 mg/ml) were added, followed by another incubation cycle at 56°C for 15–30 min. After vortexing again, the samples were placed in boiling water for 8 min and finally centrifuged at 13,000 rpm for five minutes. Prior to using a given sample as a template for PCR amplification, the final spinning step was repeated in order to separate heavy contaminants from the DNA. An aliquot standard of 2 μl of sample was used in each amplification reaction. Once extracted, the samples were stored at +4°C.

A stretch of 1100 bp of mtDNA (nucleotide positions 15971 to 00484, encompassing HVSI and 2) was then amplified using the primer pairs LF1 (5’-TTA ACT CCA CCA TTA GCA CC-3’) and LF4 (5’-TGA GAT TAG TAG TAT GGG AG-3’) whenever possible (Forster et al. 2002). The following amplification protocol was used:

| 2.00μl | template |
| 11.25μl | Distilled H$_2$O |
| 2.50μl | Yellow Sub® |
| 2.50μl | Reaction buffer (x10) |
| 2.00μl | dNTPs (x10) |
| 1.50μl | MgCl$_2$ (50mM) |
| 1.00μl | Bovine Serum Albumen (BSA x100) |
| 1.00μl | Forward primer (LF1, 10pmol) |
| 1.00μl | Reverse primer (LF2/3/4/5, 10pmol) |
| 0.25μl | Bioline Taq |

| 25.00μl | Total reaction volume |

The programme specifications for the thermocycler (Eppendorf Mastercycler Gradient®) were: 94°C for 50 seconds, then 32 cycles of 94°C for 20 seconds, 56°C for 12 seconds, 72°C for 90 seconds and a final extension of 72°C for ten minutes. The samples were then stored at between +4° and +10°C.

Amplification of this relatively long stretch often did not succeed, especially with the samples collected on Als, the collection history for which was not known in any great detail and which had been stored for c. two years prior to extraction. Direct sunlight and a number of other agents are known to cause DNA degradation over time and it is possible that these “stale” samples exhibited the first effects of such break-down. While short-term storage of buccal swab samples at room temperature has been shown
not to compromise the quality of extracts (Quinque et al. 2006), no quantitative data are available on longer-term storage. When necessary therefore, LF4 was replaced with LF5 (5'-GTT ATG ATG TCT GTT TGG AA-3'), or on occasions LF2 (5'-GAG GAT GGT GGT CAA GGG AC-3') and LF3 (5'-CAC CCT ATT AAC CAC TCA CG-3'), each of which targets an increasingly shorter sequence stretch in conjunction with LF1. In order to increase yield and specificity during amplification, the gel-loading buffer Yellow Sub* (Geneo Bioproducts, Hamburg) was employed (Haack & Vizuete-Forster 2000). The PCR products were run out on agarose gels (1–3% concentration) and purified using the QIAquick® Gel Extraction and Purification kits (Qiagen, Hilden) when needed. The benefits of gel purification and extraction were found wanting as compared with the loss of sample during the process and the price of the kits. Most commonly, therefore, amplifications that yielded only very little product were instead re-amplified and the product used directly for the sequencing reaction.

The sequencing reaction itself was carried out with the Perkin–Elmer Big Dye Terminator kit using the primers LF1–LF4 or LF1–LF5 as appropriate. The amount of PCR product added varied according to the amplification yield. 2 μl of Big Dye mix, 1 μl of forward or reverse primer (10 pmol) and enough H2O was added for a total reaction volume of 20 μl. The samples were then processed in the thermocycler in 30 cycles of 96°C for 15 seconds, 50°C for five seconds and 60°C for 120 seconds and stored at 4°-10°C, but an effort was made to process the samples as quickly as possible once they had been amplified in order to minimize the loss of sample through DNA breakdown. The sequences were read first on an Applied Biosystems Genetic Analyser 310® and later on an Applied Biosystems Genetic Analyser 3100®. Upgrading to the newer sequencing machine noticeably enhanced the sequence quality. All the sequences included in the final analysis were sequenced at least once in a forward direction using LF1 and in the reverse direction using LF4 or LF5 or twice forward in independent amplifications. Errors in human mtDNA samples are rife, even in relatively recent studies (Bandelt & Kivisild 2006), and it has been recommended that significant efforts should be undertaken to minimize the introduction of sequencing errors into mtDNA databases (Bandelt et al. 2001b; Bandelt et al. 2002; Forster 2003; Helgason & Steffánsson 2003; Herrnstadt et al. 2003; Salas et al. 2005). Thus all the sequences were checked manually for quality and obvious reading errors with the help of the Chromas® software and then automatically aligned using the DNA Alignment® software package (Fluxus Technology Ltd.). Once aligned, the individual forward and reverse sequences were concatenated and the entire sample was analysed using the mtRadius® software (Röhl et al. 2001).

Results and discussion

MtRadius analysis

On the whole, the mitochondrial haplogroup diversity observed in Danes in this study (Table 1) fits into the overall Scandinavian pattern (Helgason et al. 2001; Mikkelsen et al. 2010). Like other European populations, the Danes show a general western European affinity thought to be related to the bottleneck during the last Ice Age and the expansion after it (Forster 2004). The mtRadius® software is an analysis and visualization tool for sequence data which calculates the geographical Centre of Gravity (COG) for a given individual's phylogeographic signature in relation to a large geo-referenced database of mtDNA sequences (Röhl et al. 2001). This approach has already been employed successfully for examining the dispersal of Scandinavian Norse/Vi-

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Number</th>
<th>Per cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>H</td>
<td>89</td>
<td>47.1</td>
</tr>
<tr>
<td>H8</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>I</td>
<td>6</td>
<td>3.2</td>
</tr>
<tr>
<td>J</td>
<td>18</td>
<td>9.5</td>
</tr>
<tr>
<td>K</td>
<td>12</td>
<td>6.3</td>
</tr>
<tr>
<td>L3e</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>N1a</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>pre-V</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>T</td>
<td>11</td>
<td>5.8</td>
</tr>
<tr>
<td>T1</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>U</td>
<td>6</td>
<td>3.2</td>
</tr>
<tr>
<td>U2</td>
<td>3</td>
<td>1.6</td>
</tr>
<tr>
<td>U3</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>U4</td>
<td>8</td>
<td>4.2</td>
</tr>
<tr>
<td>U5</td>
<td>5</td>
<td>2.6</td>
</tr>
<tr>
<td>U8</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>V</td>
<td>12</td>
<td>6.3</td>
</tr>
<tr>
<td>W</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Undetermined</td>
<td>5</td>
<td>2.6</td>
</tr>
<tr>
<td>TOTAL</td>
<td>189</td>
<td>100.0</td>
</tr>
</tbody>
</table>
MtDNA and the pioneer colonization of Scandinavia

King populations across Atlantic Europe (Forster et al. 2004). MtRadius® offers phylogeographic specificity by first providing an output for each individual analysed, consisting of a map showing the distribution of identical or near-identical sequences. These individuals are shown as dots on a map with the size of the dot corresponding to the number of individuals at the given location and the distance of the dots from the calculated COG indicated by their colour: black <400 km, grey 400–800 km, and white >800 km. The latter category is considered uninformative as these haplotypes show very wide, non-specific distributions, i.e. this category reflects primarily very ancient distribution patterns on geographical scales very much larger than the one of interest here. A list detailing the identical and near-identical database entries is also provided for each individual. Finally, mtRadius® produces a summary map for the entire sample.

The present sample was supplemented with 33 Danish sequences reported in Richards et al. (1996) and compared with a database of over 25,000 mitochondrial DNA sequences with a global (but primarily European) distribution (Röhl et al. 2001). The results are shown in Fig. 2. Considering the entire Danish sample, i.e. all the black, grey and white dots, a general and unsurprising European affinity is evident, but when considering only the more informative grey and black dots, a western affinity can clearly be seen. We therefore conclude that the mtDNA sequence data provide strong support for H2, i.e. a western dispersal trajectory.

Genetic studies on northern Scandinavian populations support this western European affinity. Rootsi et al. (2004) argue that the high frequency of Y-chromosome haplotype I1a reflects a Late Glacial western pattern of dispersal, while Passarino et al. (2002) further link northern Scandinavian mtDNA to western Europe, but also, interestingly, see a significant male input from eastern Europe in the form of Eu19/R1a1/M17 (YCC 2002). While there are important differences in the scales of resolution between the maternal and paternal line signatures (Forster et al. 2004), these differences in signature may reflect genuine sex-specific demographic histories during the colonization of Scandinavia. Both Passarino et al. (2002) and Karlsson et al. (2006) argue that the Y-chromosome patterns are most consistent with a central European source population and link this colonization with the Ahrensburgian culture, which – although ultimately derived from western source populations – is
characterized by a north-central European distribution (see Cziesla 2007). In the light of demographic considerations and with the support of archaeological data, it has recently been argued (Riede 2007, 2009) that the very first dispersals of human populations into Scandinavia, those associated with the Hamburgian culture, were in fact unsuccessful. Instead, the demographically viable establishment of hunter-gatherer-fisher groups in Scandinavia is seen to be related to the Ahrensburgian in north-central Europe (Schmitt 1999) and was facilitated by the dramatic increase in terrestrial and marine productivity after the end of the Last Ice Age (Schmitt et al. 2006; Schmitt et al. 2009). Differential mortality and divergent regimes of reproductive success amongst the male and female pioneer colonizers of Scandinavia could have produced a pattern of western affinity for the maternal signature (perhaps associated with the Hamburgian culture), and a central European component in the paternal signature (perhaps associated with the Ahrensburgian northward migration).

Rare haplotypes

Although our Danish sample sits firmly within its European genetic context, as shown above, the dominance of a few common haplotypes is complemented by haplogroups occurring at very low frequencies, some of which are generally considered rare in present-day northern Europe. Likewise, a recently published sample of living (Caucasian) Danes (Mikkelsen et al. 2010) contains haplotypes at low frequencies, albeit in a different composition from that in our sample. Recent aDNA studies on Medieval (Rudbeck et al. 2006) and Iron Age (Melchior et al. 2008) Danish populations have similarly revealed such rare haplotypes. Interestingly, the samples from the different time-slices show somewhat different assemblies of such rare haplotypes (Table 2).

The “surprising diversity” observed in the ancient (early Christian) Danish mtDNA pool by Rudbeck et al. (2006:428–9) can be traced from at least the Iron Age onwards. While one can only speculate on the reasons for the occurrence of rare haplotypes in historical samples, the family histories collected alongside the genetic samples in this study allows a better resolution, albeit not for all the individuals representing rare haplotypes. The individual with haplotype L3e (in the Panum sub-sample), for instance, has familial connections with the West Indies, where African haplogroups such as L3e are common (Bandelt et al. 2001a). The sample recently presented by Mikkelsen et al. (2010) contains only Caucasian Danes, but no information was collected on their geographical ancestry. The difference in rare haplotypes between this sample and our own may at least partly be attributed to this methodological difference. These data aptly reflect the constant, on-going turn-over of rare haplotypes in a given population, whilst the aDNA studies give temporal depth to this process.

Conclusions

Southern Scandinavia played an important corridor role in the re-colonization of northern Europe after the Last Ice Age. On the basis of a phylogeographic analysis of mtDNA using the mtRadius® software we favour a primary western dispersal trajectory. If this pattern reflects Palaeolithic population movements, it may be more parsimoniously linked to the Hamburgian culture. This would stand in some contrast to current interpretations of the NRY-chromosome data, which, owing to considerable central European affinities, is interpreted as reflecting human dispersal at the very beginning of the Holocene warm period as reflected in the material culture of the Ahrensburgian techno-complex (Passarino et al. 2004). In order to reconcile these competing interpretations, we suggest that more complex demographic scenarios, including sex-specific bottlenecks and multiple secondary dispersals prior to the Neolithic – for example those that occurred in
the wake of the Laacher See volcanic eruption (Riede, 2008) – might have to be invoked. Taken together, both mitochondrial and Y-chromosome data from Denmark strongly favour H2, a western direction of dispersal. It is less clear, however, whether these data can confidently be linked to the dispersal of Late Glacial hunter-gatherer populations at all, or whether later dispersals were responsible.

We know from aDNA (Haak et al. 2005; Burger et al. 2007; Bramanti et al. 2009; Malmström et al. 2009) and from archaeological evidence (Shennan & Edinborough 2007) that the population history of northern Europe did not stabilize even after the establishment of agriculturally-based economies, and that a series of dramatic climatic and environmental changes took place around the end of the Last Ice Age (Riede 2008, 2009). As Bramanti et al. (2009:139) have pointed out, “the extent to which modern Europeans are descended from incoming farmers, their hunter-gatherer forerunners, or later incoming groups remains unresolved”. The settlement history of southern Scandinavia has clearly been complex and this is reflected in particular in the changing composition of the rare haplotypes found in ancient and living populations in the region. We therefore urge caution in interpreting the population genetic data available at present. Nonetheless, the current sample has assisted us in evaluating the opposing hypotheses for the pioneer human re-colonization of Scandinavia and it allows us to place aDNA studies of historical populations in a more robust context. Thus the analysis of population genetic data in the context of specific issues in history or prehistory can lead to many new insights (see Borglum et al. 2007). Indeed, recent advances in sequencing technology hold the promise of obtaining very detailed biological knowledge about past individuals and populations (Rasmussen et al. 2010), complementing the insights into past social and material aspects obtained from traditional archaeological, anthropological and historical sources.

Acknowledgements

We acknowledge the financial support of Cambridge University (McDonald Institute of Archaeological Research, Pembroke College, Kurt Hahn Trust), where the research for this paper was carried out. We particularly thank Peter Forster for laboratory support and guidance and for his assistance with the phylogeographic analysis. The comments of two reviewers improved the quality of the manuscript markedly. All the remaining mistakes are entirely the authors’ responsibility.

English language revision by Malcolm Hicks.


