The Neolithic transition in Europe: archaeological models and genetic evidence

Martin Richards
Department of Chemical and Biological Sciences, University of Huddersfield, UK
m.b.richards@hud.ac.uk

ABSTRACT – The major pattern in the European gene pool is a southeast-northwest frequency gradient of classic genetic markers such as blood groups, which population geneticists initially attributed to the demographic impact of Neolithic farmers dispersing from the Near East. Molecular genetics has enriched this picture, with analyses of mitochondrial DNA and the Y chromosome allowing a more detailed exploration of alternative models for the spread of the Neolithic into Europe. This paper considers a range of possible models in the light of the detailed information now emerging from genetic studies.

IZVELEČEK – Glavni vzorec evropskega genskega bazena je gradient klasičnih genskih markerjev v smeri jugovzhod-severozahod. Tak marker je na primer krvna skupina. Njen gradient so populaciji-sk genetiki prvotno pripisovali demografskemu vplivu neolitskih kmetovalcev, ki so se razširili iz Bliznjega vzhoda. Molekularni genetiki so to sliko obogatili z analizami mitohondrijske DNA in Y kromosoma, kar je omogočilo podrobnejši razvoj alternativnih modelov razširjanja neolitika v Evropo. V članku pretehtamo več možnih modelov v luči podrobnih informacij, ki jih danes dajejo genske raziskave.

KEY WORDS – Neolithic farmers; Mesolithic foragers; mitochondrial DNA; Y chromosome; phylogeography

MODELS OF NEOLITHIC DISPERSAL

How did the Neolithic spread from the Near East into Europe? In the past, this issue has often been polarised as an either/or between ‘demic diffusion’, usually taken to mean a large-scale movement into Europe of Near Eastern farming people, versus ‘cultural diffusion’, in which it is rather the idea of farming that spreads. However, in recent years the range of possible models has become rather more nuanced. Zvelebil (2000) has listed seven possible mechanisms:

1. Folk migration. This is the traditional migrationist explanation: the directional movement of a whole population from one region to another, leading to genetic replacement.
2. Demic diffusion, by means of a wave of advance.
3. Elite dominance, in which a social elite penetrates an area and imposes a new culture on the local population.
4. Infiltration of a community, for example by small numbers of specialists fulfilling a particular need, such as livestock farmers.
5. Leapfrog colonization by small groups targeting optimal areas, to form an enclave surrounded by indigenous inhabitants.
6. Frontier mobility, or exchange between farmers and foragers at agricultural frontier zones;
7. Regional contact, involving trade and exchange of ideas.

In this article, we will ask whether it is possible to use the existing genetic evidence to begin to distinguish these possibilities.

What would be the genetic predictions for each of these models? If we assume, for the sake of argument, that the Near East and Europe can be cleanly...
partitioned and were genetically distinct prior to the onset of the Neolithic, then different models may be taken to predict different genetic patterns.

The first model is classic “migrationism” and would involve genetic replacement, so that the sink region (Europe) should be genetically indistinguishable from the source (the Near East), except for any differentiation that had taken place within the last 8000 years. Model (7) would involve no movement of genes whatsoever – Ammerman’s “indigenism” (Ammerman 1989). This would include both cultural diffusion (Dennell 1983; Barker 1985; Whittle 1996) and separate development, in which the social and ideological, rather than economic, aspects of the Neolithic take centre stage (Hodder 1990; Thomas 1996; 1998). In this case, the source and sink regions should remain genetically distinct, except for the effects of any post-Neolithic gene flow between them.

Models (2) to (6) are all “integrationist” (Zvelebil 2000) in character, involving both the arrival of new genetic lineages in an area, and the eventual acculturation of the indigenous communities. Elite dominance might show minor evidence of newcomers, although it might not be relevant to the question of the early Neolithic (Renfrew 1987). The wave of advance model predicts continent-wide genetic clines (Ammerman and Cavalli-Sforza 1984). Infiltration and leapfrog colonization would be likely to leave traces of Near Eastern lineages in the regions where they had occurred, but in patches rather than in the form of clear clines. Frontier mobility would allow for genetic exchange between colonised, newly Neolithic areas such as central Europe, and forager strongholds to the north and west. In each of these, however, any genetic discontinuities might tend to be eroded over time as the effects of subsequent gene flow acted to blur the picture.

CLASSICAL MARKERS

It has long been assumed (by population geneticists, at least), that classical markers support the Ammerman and Cavalli-Sforza (1984) model of demic diffusion by means of a wave of advance. This model depended on a view of the early Neolithic that emphasized sedentism, local population growth, and expansion into more marginal environments. Ammerman and Cavalli-Sforza (1984) modelled the expansion using Fisher’s “wave of advance”, and compared the results with radiocarbon maps of the spread of the “Neolithic package” across Europe. The “package” included emmer wheat, einkorn wheat and barley – whose wild progenitors occurred only in the Fertile Crescent region of the Near East – domestic animals, pottery, ground and polished stone tools, and houses. However, they often relied upon one or two “marker” items, rather than the whole package.

This smoothing led to their estimation of a uniform rate of spread across Europe of about one kilometre per year, or 25 kilometres per generation – from Greece to the British Isles in about 2500 years. This led them to the idea of a single all-embracing mechanism, which they called “demic diffusion”. This was intended to be distinct not only from cultural diffusion, or the spread of ideas, but also from good old-fashioned directed colonization. The mechanism they proposed was the wave of advance: logistic population growth (resulting from agricultural surpluses and storage) plus random local migratory diffusion or range expansion. They referred to it as “colonization without colonists”.

The so-called classical markers, or non-DNA markers, comprise allele frequencies for blood groups, the tissue antigen HLA system, and some enzymes. The signal from these markers was not strong, and moreover, different markers gave different signals. Furthermore, it was clear that Europe and the Near East were not as genetically differentiated as Ammerman and Cavalli-Sforza would have liked. So they took a multivariate approach, choosing principal-component (PC) analysis (Menozzi et al. 1978), and presented the results, component by component, as synthetic contour maps, showing the changes in frequency with geography.

The first PC, accounting for about 27% of the total variation in classical marker frequencies across Europe and the Near East, famously showed a gradient from the southeast to the northwest, with the Near East at one pole and Europe at the other. This pattern was clearly reminiscent of the radiocarbon map for the spread of the Neolithic. This was, Cavalli-Sforza and his colleagues believed, strong evidence for a mixed demic diffusion hypothesis, in which there was both a demic expansion and intermarriage with local hunter-gatherers on the way. The second and third components (explaining about 22% and 11% of the variation respectively) showed gradients that were oriented roughly southwest-northeast and east-west. Because of their lower impact on the genetic variation, they were assumed to have been the result of processes that had taken place since than the Neolithic.
The conclusions of Ammerman and Cavalli-Sforza and their colleagues were supported by Sokal and colleagues (Sokal et al. 1989; 1991), using spatial autocorrelation analysis. This approach also indicated that about a third of classical markers were arranged in a southeast-northwest cline. With this backing, the assumed model of surplus-driven population growth and expansion gained ground and began to be taken for granted amongst population geneticists. Despite the inability of these methods to quantify the demographic impact of the Neolithic newcomers, the role of the putative pioneers came to be emphasized at the expense of the indigenous Mesolithic peoples of Europe. Furthermore, the idea that the PC maps could be interpreted chronologically, like archaeological stratigraphy, also took hold (Cavalli-Sforza 1996).

However, gradually some criticisms were expressed. Why interpret the first PC solely in terms of Neolithic expansion? Europe is a small peninsula of the Eurasian landmass, and as such is likely to have been the sink for many dispersals throughout prehistory. The PC maps were much more likely to represent a palimpsest of dispersals, each one overwriting the last (Zvelebil 1989; 1998). The idea of “one PC—one migration”, suggested quite specifically by Cavalli-Sforza, was highly implausible; and this disposed equally of the idea that principal components provided a genetic stratigraphy. Indeed, the problematic second PC, running southwest–northeast, was increasingly looking as if it might be explained at least in part by Lateglacial hunter-gatherer expansions, preceding the Neolithic by more than 5000 years (Torroni et al. 1998).

The archaeological aspects of Ammerman and Cavalli-Sforza’s work also sustained criticism. Items in the “Neolithic package”, it was pointed out, rarely moved together, except in southeast and central Europe, and they were often exchanged into Mesolithic communities (Thomas 1996; Zvelebil 1986; Price 2000). This could have led Ammerman and Cavalli-Sforza to over-estimate the impact of the Neolithic and the uniformity of its spread. More recent studies have tended to emphasize that the spread of the Neolithic was a heterogeneous process, with no evidence in the archaeological record for large-scale continent-wide immigration (Pluciennik 1998; Zvelebil 2000). Furthermore, the link between Neolithic populations and high population density, and Mesolithic ones and low density, has not survived more detailed study. The archaeological and palynological records suggested that the high growth potential of Neolithic communities was very unlikely ever to have been achieved during the early millennia of farming (Willis and Bennett 1994; van Andel and Runnels 1995; Roberts 1998.154–8). At the same time, riverine and coastal Mesolithic communities may well have allowed the growth of affluent, complex foraging communities, with higher population densities, and a much higher degree of sedentism, than once assumed (Zvelebil 1986).

Molecular markers and phylogeography

In the 1980s, it became possible to analyse not merely the products of certain genes, as had been done in the “classical” analyses, but the DNA sequences of the genes themselves. For studies of evolution and migration, attention has focused on the two non-recombining genetic loci in humans. The mitochondrial DNA (mtDNA) is present in both sexes, but inherited only down the maternal line, whereas the Y chromosome is present only in males and is inherited only from father to son. Although future studies will focus on the remaining, recombining parts of the genome – the X chromosome and the autosomes – there are two particular advantages to the non-recombining systems, in which variation is not reshuffled between different lineages with every passing generation, but is inherited down a single line of descent.

Phylogenies, or genealogical trees, can be estimated. Both mtDNA and the Y chromosome can be seen as genetic systems in which mutations fall onto an independently-formed genealogy: the maternal and paternal lines of descent, respectively. Any sample of individual subjects will have a defined set of genealogical relations on both the maternal and paternal side, so that in principle a tree of ancestry could be reconstructed for each. The mtDNA and the Y chromosome both allow us to estimate those trees, because both systems have recorded a trace of the pattern of descent, as mutations have inscribed variants into their DNA sequences during the course of history. This implies a dramatic increase in the resolution of processes involving individuals, such as prehistoric dispersals (Richards and Macaulay 2000).

Lineages can be dated, using the molecular clock. Although not as reliable as radiocarbon dating, this represents a great improvement on the analysis of frequencies of classical markers where, as we have seen, dating is a problem even if it could be assumed that a particular genetic pattern has been produced by a single process.
These developments have led to the development of what has been termed the “phylogeographic” approach (Richards et al. 1997; Bandelt et al. 2002). Phylogeography is a heuristic tool for interpreting complex population-genetic data that tries to make maximum use of reconstructed trees of descent, along with the geographic distribution and diversity of genealogical lineages; it is effectively the mapping of gene genealogies in time and space (Ariese 2000). The process of testing phylogeographic hypotheses always entails making assumptions, and inevitably has to be carried out within a model or framework based on external information (such as from archaeology). Even so, the assumptions themselves can often be susceptible to empirical investigation, and may often be less unrealistic than those of more traditional population-genetics approaches (Richards et al. 2000).

MITOCHONDRIAL DNA

The first major application of phylogeographic procedures to the question of European genetic variation was an analysis of mitochondrial DNA (mtDNA) (Richards et al. 1996). This work made use of a new phylogenetic-network approach to tree reconstruction, developing new phylogeographic approaches, such as founder analysis, to the study of migration and colonization.

Founder analysis works by comparing the genetic variation in a region that has been settled (the sink population) with that in likely source populations, in order to identify founder sequence types and use them to date individual migration events (Richards and Macaulay 2000). This is done explicitly to avoid the charge that “the age of a population is not the age of the common molecular ancestor of its set of DNA sequences”, although curiously this criticism continues to be made (Barbujani et al. 1998; Chikhi et al. 1998; Barbujani and Chikhi 2000). When there is an individual migration event from the source to the sink region, so that a founder event occurs, the molecular clock is effectively reset, so that the descendants of that individual can be regarded as members of a new line of descent tracing to the time of arrival. The molecular age of the founder type in the source population will of course be older – perhaps much older. Founder analysis proceeds by subtracting from the mutational variation in the sink population that fraction of the variation that arose in the source population and has been carried into the sink region by the founders during the colonization process. This is done so that only the new mutations that have arisen since the colonization are used when estimating dates.

The initial, rather tentative, results from European mtDNA suggested that the majority of lineages appeared to descend from founders of Middle or Late Upper Palaeolithic origin, implying re-expansions in the Lateglacial or post-glacial period. Only a fifth or less dated to the Neolithic (Richards et al. 1996; 1998).

Further work by Torroni and colleagues (1998; 2001) strikingly confirmed the existence of major Lateglacial expansions from southwest Europe, suggesting a plausible explanation for the second PC of classical markers. Meanwhile, Richards et al. (2000) carried out a much more thorough founder analysis of a greatly enlarged Near Eastern and European mtDNA data set. Although it is very difficult to extrapolate to the scale of the immigration at the time, it is possible at least to estimate the proportion of lineages in the modern population that descend from one or other immigration event. They found that about three-quarters of modern mtDNA lineages could be traced to just eleven ancestors (the remaining quarter comprising a larger set of minor founders). Under a range of assumptions, the putative Neolithic component in modern Europe (i.e., those lineages that appeared from the Near East about 9000 years ago) occurs at between 12–23%, the best estimate being ~13%. Lateglacial expansions were conflated with preceding Middle Upper Palaeolithic immigration, but between them accounted for almost 70% of modern lineages. It appeared that, on the maternal line of descent, only a small fraction of modern Europeans were descended from Near Eastern farmers; in the main, they were descended from indigenous European foragers, who adopted farming later on.

A number of critiques of this work have appeared, guided by classical population-genetics approaches rather than phylogeography, in particular the dating of “population splits” (Chikhi et al. 1998; Barbujani and Bertorelle 2001). This approach, however, fails to provide dates that are genuinely meaningful in terms of demographic history (Bandelt et al. 2002). Critiques of the statistical validity of the founder analysis may have more force, since it relies on the sample size in the source population being adequate to identify all of the most important founder types. However, some limited resampling tests have given very similar results, particularly for the Neolithic contribution (Richards et al. 2000). This reanalysis used only the “core” Fertile Crescent data, omitting
Anatolia, Egypt, and the southern Caucasus. It may also help, therefore, to address the Eurocentric bias of the main analyses, which draw rather a sharp division between “Europe” and the “Near East” at the Bosporus and Caucasus mountains (M. Özdoğan, personal communication).

Richards et al. (2000) also repeated the analysis at the regional level. It must be pointed out that this approach has serious limitations. In the first place, the results for any one region are based on fewer data and are therefore naturally associated with greater uncertainty. Moreover, the regional data are of variable quality, and may poorly represent the deep ancestry of lineages within each region in some cases (such as eastern Europe and Greece). Finally, the results are, at best, estimating the proportion of lineages in the present-day population that can be attributed to each founder event from the Near East (or to bottlenecks within Europe), rather than from the immediate source region. Given these caveats, the results may nevertheless bear some discussion.

The analysis suggested that the highest Neolithic impact was on southeast Europe, central Europe, northwest and northeast Europe, which showed values of 15–22% Neolithic lineages each. The Neolithic lineages are mainly from haplogroup J, and include a specific subset of J lineages, called J1a, that are largely restricted to this region and seem to be a marker for the Linienbandkeramische Kultur (LBK) and post-LBK dispersals (Richards et al. 1996). For southeast and central Europe, a relatively high Neolithic component seems congruent with the usual interpretation of the archaeological record. There is some consensus that the Balkan Neolithic and the central European LBK were the result of direct colonization, although there is debate about the extent of acculturation along the way (cf. Gronenborn 1999, 2003; Tringham 2000; Budja 2001). Acculturation may indeed have taken place in between the two processes, where there was a substantial break in the expansion (Bogucki 2000; Zvelebil 2000). The mtDNA results suggest that colonization from (ultimately) the Near East did indeed take place, and that the descendants of Near Eastern colonists are represented in the central European populations of the present day. Nevertheless, more than three-quarters of the surviving lineages are the result of acculturation of indigenous foraging peoples. This appears to broadly support “integrationist” models (Zvelebil 2000; 2001), such as pioneer “leapfrog” colonization (directed towards suitable land) and acculturation and genetic exchange across the agricultural frontier during the phase in which aspects of farming become available to the surrounding foraging populations. Strontium isotope analysis has recently suggested immigrations of non-local people into LBK settlements from very early times (Bentley et al. 2002). It is possible that some of these were brought in from the surrounding foraging communities (Gronenborn 1999).

The presence of Near Eastern lineages at similar frequencies in the northwest seems less consistent with Zvelebil’s model, which suggests that a long-term frontier was established on the north European plain, and that the transition to farming to the north, northwest, northeast and southwest took place largely by acculturation. However, it also conflicts with the patterns of the classical markers and the Y chromosome (see below), in which the putative “Neolithic” lineages or alleles tend to zero towards the north-west periphery of the continent. If we take the mtDNA patterns seriously, perhaps there were female-only exchanges between the post-LBK peoples of the North European plain and the northwest across the agricultural frontier (Wilson et al. 2001). Alternatively, there may have been acculturation at the LBK frontier, after which predominantly Near Eastern mtDNAs, but predominantly acculturated Y chromosomes (by chance in both instances) moved northwest (Renfrew 2001). It is also, of course, possible that the mtDNA lineages were dispersed into the northwest by later dispersals.

There are fewer Neolithic-derived mtDNA lineages along the Mediterranean and the Atlantic west (about 10%). The sample from the eastern Mediterranean is small and not well provenanced, but the results nevertheless appear compatible with the maritime colonization of Greece by Near Eastern pioneer groups (Perlès 2001). As in central Europe, some of the putative Neolithic lineages further west are again regionally specific: for example haplogroup J1b, which appears to have leap-frogged from the Near East straight across to the Atlantic façade. This certainly seems consistent with the archaeological view of maritime colonization in the west alongside acculturation of quite dense, sedentary Mesolithic communities (Barnett 2000; Zilhão 2000; 2001).

**THE Y CHROMOSOME**

Unlike the mtDNA work on the Neolithic transition, the first major publication on the Y chromosome (Semino et al. 2000) had been prefigured by earlier studies that had already identified a demic compo-
These reconstructions are made on the basis of the mtDNA work. They identified several potential Neolithic markers that implied a Near Eastern Neolithic contribution to Europe as a whole of less than 25%. There have been recent criticisms of their interpretation by Chikhi et al. (2002), on the grounds that an admixture approach suggests a much higher putative Neolithic contribution than the crude estimates. However, their arguments are unconvincing, since an admixture approach seems quite inappropriate in the context of the questions under consideration, and suffers from some of the weaknesses of the classical approach (such as lack of dating).

It is noticeable, though, that the putative Neolithic lineages are markedly more common along the Mediterranean than in central Europe, which contrasts somewhat with the mtDNA picture described above. Without a founder analysis, such as has been done for mtDNA, it is certainly likely that earlier and later processes may be conflated: the palimpsest problem again. The question is to what extent. King and Underhill (2002) have argued that the high correlation between the distribution of painted pottery and anthropomorphic clay figurines and some of the putatively Neolithic Y chromosomes indicates that indeed at least some of the latter do represent early Neolithic settlement. This implies that, on the male side, intrusive lineages from the Near East only spread through the first burst of Neolithic settlement in Europe around the eastern Mediterranean basin, but were not carried to an appreciable extent into central Europe with the LBK. This in turn supports the view that high levels of acculturation took place in the Balkans prior to the LBK expansion (Gronenborn 1999; 2003). The Near Eastern lineages that spread through the eastern and central Mediterranean in the early Neolithic would have been subsequently overlaid by later Near Eastern dispersals. It is also possible that Neolithic colonisation of the Mediterranean from the Near East involved maritime pioneers who were predominantly male, and that this goes some way to explaining the much higher male contribution of Neolithic lineages in the east and central Mediterranean (Perlès 2001).

CONCLUSIONS

Nothing intrinsically associates any particular mtDNA, or Y chromosome, with the spread of the Neolithic. These reconstructions are made on the basis of the estimated time of arrival of particular lineages and their geographical distribution. Alternative explanations of the same patterns are inevitably possible, depending on the breadth of possible frameworks made available by the archaeological evidence (Bandelt et al. 2002). But given these caveats, what can be suggested about the process of the Neolithisation of Europe from the study of the genetics of modern European populations?

All of these marker systems suggest that there was indeed a process of colonization during the spread of the early Neolithic into central and western Europe. This rules out, as decisively as is likely to be possible with genetic evidence, models based solely on cultural diffusion and acculturation, or separate development (model 7). This pattern would also seem to rule out elite dominance (model 3).

At the other extreme, the mtDNA and Y-chromosome evidence both imply a minor overall contribution to modern lineages of less than a quarter, suggesting that large-scale demic diffusion (model 2) or even replacement (model 1) can also be ruled out. However, discriminating smaller-scale demic diffusion, which could have a large cumulative impact in terms of, for example, language replacement (Renfrew 2001), is more difficult. Small-scale demic diffusion by means of a wave of advance would be expected to generate clines, which are indeed seen in some classical and some molecular markers, including the Y chromosome. In the case of classical markers, whether the clines to any extent reflect a Neolithic expansion is hard to determine; in the Y chromosome, however, it does seem that they may be partly the result of a Neolithic dispersal. However, comparison of founder and PC analyses of mtDNA imply that many components of the clines may be the result of both more ancient and more recent expansion events (Richards et al. 2002).

We are left with a number of mixed migrationist/diffusionist models that are not mutually exclusive (Gronenborn 1999; 2003; Zvelebil 2000; 2001). The evidence of mtDNA and the Y chromosome seems to be consistent with pioneer leapfrog colonization and infiltration of southeast and central Europe, and the subsequent filling acculturation of much larger numbers of indigenous foragers. There may have been a wave of advance (Zvelebil’s “starburst demic diffusion”) during the rapid expansion in the LBK area, but if so, it must have largely involved mtDNA and Y-chromosome lineages from assimilated Balkan foraging populations, rather than from the Near
East (Gronenborn 1999). Archaeological evidence is now emerging from both ceramics and lithics for the assimilation of Mesolithic groups into LBK settlements (cf. Gronenborn 2003).

There is some evidence for further colonization from the LBK zone into the northwest, including the British Isles, whereas the pattern in Scandinavia might be explained by frontier exchange. The Atlantic west seems also to have experienced distinct, presumably maritime leapfrog colonization events from the direction of the west Mediterranean coastline. The movements into the northwest seem either not to have involved men, or to have involved male lineages that had undergone acculturation, and were therefore indigenous to central Europe. In all or most regions of Europe, even in the LBK zone, there seems to have been substantial local adoption of agriculture.

ACKNOWLEDGEMENTS

I would like to thank Detlef Gronenborn for critical advice on an earlier version of this manuscript, and Mehmet Ozdoğan and Catherine Perles for valuable comments during the conference at which this work was presented.

REFERENCES


