Chapter 35

Complex Signals for Population Expansions in Europe and Beyond

Kristiina Tambets, Helle-Viivi Tolk, Toomas Kivisild, Ene Metspalu, Jüri Parik, Maere Reidla, Michael Voevoda, Larissa Damba, Marina Bermisheva, Elsa Khusnutdinova, Maria Golubenko, Vadim Stepanov, Valery Puzyrev, Esien Usanga, Pavao Rudan, Lars Beckmann & Richard Villems

François Jacob, in his brilliant 'The possible and the actual' (Jacob 1982), reminds us that 'scientific investigation begins by inventing a possible world, or a small piece of a possible world'. One may add that the space allowed for *the possible* is likely to be in strong positive correlation with the level of our ignorance. What chance, then, when discussing language / farming / gene dispersals, do we have to identify *the actual* from a plethora of *possible* scenarios?

Since the presentation of 'African Eve' (Cann *et al.* 1987; Vigilant *et al.* 1991), the last decade has demonstrated an increasingly better understanding of the phylogeny and phylogeography of mtDNA and of the Y chromosome. Here, the first influential achievement was a series of papers from Emory (reviewed in Wallace 1995) where, *inter alia*, it became obvious that human maternal lineages world-wide are very clearly structured geographically. This knowledge came thanks to phylogenetic analysis of the coding part of the mtDNA genome. Secondly, as Richards *et al.* (1996) have shown, the mtDNA hypervariable 1 (HVR 1) region offers an increased resolution of a phylogenetic tree, in particular as far as Europeans are concerned. Although mtDNA hypervariable region sequences started to accumulate in quantities (thanks largely to forensics), it soon became obvious that the results coming from RFLP analysis or the HVR sequence(s) alone were not informative enough to go further. Quite the opposite; it became clear that trees, based on HVR 1 sequence alone, were often phylogenetically wrong. However, a synthesis of what is known about polymorphisms in the coding region (extensive RFLP as a tool) and HVR (direct sequencing) removes most of the ambiguities and leads to a much better understanding of the details of the topology of the phylogenetic tree of mtDNA (e.g. Macaulay *et al.* 1999). This analysis owes much to the use of median networks as an approach (Bandelt *et al.* 1995).

In this contribution we demonstrate that coalescence age calculation of the monophyletic branches of the mtDNA phylogenetic tree, applied together with a detailed phylogeographic knowledge, is an instrument which provides new insight into demographic processes of the past and, in particular, allows to see informative differences there, where mere haplogroup frequency calculations are able only to register flat landscapes.

General

How much further can one go in resolution? It is obvious that 'the ultimate' answer lies in analyzing, in all collected samples, all 16,500 plus nucleotides of the mtDNA genome — to carry out total (high fidelity!) re-sequencing. There are now at least a thousand fully sequenced mtDNA genomes at hand and this body of data, although rather time-consuming to analyze, is very useful in 'fine-tuning' phylogenetic analysis (Richards & Macaulay 2001). However, when we speak about many thousands of samples, total re-sequencing is not yet a viable approach and would probably be unnecessary either. While extensive re-sequencing did reveal a number of new polymorphisms very useful for fine-scale analysis, and allowed the resolution of a number of ambiguities, it did not create a need to revise the basic topology of the mtDNA tree as it had been deduced in relatively fine detail already. It does not mean, though, that an additional total or partial re-sequencing of mtDNAs is not needed any more, specifically where hitherto less understood variants are concerned.

Perhaps the most valuable aspect of mtDNA diversity analysis as an 'archaeogenetic tool' lies in the possibility of estimating coalescence ages of individual lineage clusters. In saying 'valuable', it is only fair to add that there are different views on the reliability of the mtDNA clock. Although coalescence estimates for human mtDNA lineage clusters (haplogroups) are often presented with rather large standard deviations and are prone to possible systematic errors in special circumstances, the approach as such is a tool to be polished further, not discarded.

One complicated question, though, is whether the diversity within a particular clade which we observe at present, among a population or a group of populations in a contiguous area, has arisen *in situ* or was, at least partially, already present among some ancestral population and then carried to new places by a large enough number of people to keep the pre-existing diversity 'alive'. This is a valid question universally (e.g. compare the coalescence ages of Amerindian mtDNA lineage clusters of more than 20,000 years with the much younger archaeological evidence for the peopling of the Americas), and to give a satisfactory answer is usually not easy. A thorough analysis of the phylogeography of individual lineage clusters with a reliable identification of founder haplotypes may help here.

It is justifiable to expect that the main driving forces behind the ancient demographic behaviour of human populations were more or less directly related to the availability of food and hunting and gathering territories. Consequently, one may presume that pre-Neolithic population expansions, on a Eurasian scale, could be largely attributed and traced to periods when large 'virgin' land areas first became available. The first colonization of Eurasia by modern humans and, particularly for Europe, the postglacial re-colonization of formerly glaciated territories, are two obvious examples. Less discussed in the western Eurasian context are other possible re-colonizations, particularly of lands deserted because of the extreme aridity that accompanied the LGM and made large areas of northern Africa and the Near and Middle East inhospitable for humans. And then there was the Younger Dryas, with its very abrupt return to cold and dry conditions, possibly even more dangerous for human survival than equally harsh but much slower environmental changes.

That food production (agriculture) allows for higher population densities than hunting and gathering certainly seems to be self-evident. However, the historic context is always concrete and such general statements should be weighed against specific archaeological evidence (Bellwood 2001; Renfrew 2000). As far as Neolithic Europe is concerned, the literature addressing the influence of Neolithic Anatolian/Near Eastern genes in the extant gene pool of Europeans is already extensive and widely known. We refer here only to Richards *et al.* (2000) where, for the first time, a more complex model of gene flow between the Near East and Europe was considered in the interpretation of the phylogeography of mtDNA lineages. This model assumes reciprocal movements of people and is probably more realistic than one assuming only a one-way flow.

Modern humans started to colonize Europe about 40,000-50,000 years ago. Population density underwent many profound changes. How these phases of expansion/stabilization and regression/re-occupation may have influenced the linguistic situation is discussed by Renfrew (2000). He pays attention mostly to the more recent periods, whereas we will try to start 'from the beginning'. In the Eurasian context, it would be possible to consider the first expansion phase (i.e. the first colonization of Eurasia by modern humans) as having generated the first language spread zone. Next, one may ask how a drastic shrinkage of the area inhabited during the LGM (particularly in Europe) influenced the linguistic situation. While several much shorter post-LGM cold phases like the Younger Dryas arose and disappeared abruptly, within perhaps a few generations (van Andel 2000), it seems that the LGM itself developed in cold intensity much more slowly — a process that might well have allowed time for 'an organized retreat' into refugia. What proportion of the pre-LGM mtDNA and Y-chromosomal diversity survived in these refugia? And what was happening with languages?

However remote the pre-LGM period may seem to us — but notice a recent confirmation of sophisticated Aurignacian cave art in France (Valladas *el al.* 2001) — it lasted for more than 15,000 years and was a period when some of the most remarkable pan-European cultures like the Aurignacian and

Gravettian flourished. Could the following glacial maximum period (roughly 24,000-16,000 BP) represent a strong convergence phase for the pre-LGM languages in refuge areas, coupled with a simultaneous and profound divergence enforced by the geographic isolation of the principal refugia? And how complete was this isolation?

For example, was the 'Periglacial refugium' (Dolukhanov 2000) in Eastern Europe indeed isolated for millennia from other (likely) refugia in France and Iberia, in southern Siberia, in the Balkans, and along the eastern Black Sea coast? And were the latter two isolated from eastern Anatolia and the Levant, or was there in fact traffic, the migration of humans, forth and back? What influence would such traffic have had on the then-existing patterns of languages? Is this question at all relevant for the emergence of language families as we know them at present? Is it beyond knowability? If so, then archaeology and genetics can operate free of linguistic constraints over the time frame under discussion. Is it meaningful to explore, just for the sake of curiosity, an alternative working hypothesis according to which (some) language families which are currently spread across western Eurasia may have had their origins in a convergence phase during the LGM, where refugia brought together languages which earlier, during perhaps the preceding 15,000 years (or even much longer), were subject to a 'spread zone' scenario?

It may well be that the conventional methods of historical linguistics and glottochronology do not normally allow us to penetrate deeper than, say, 5000-8000 years. Sometimes it seems that this chronological barrier is understood to mean that a particular language family indeed arose at this time depth, while it may be just a time-line beyond which further reconstruction is impossible. Therefore, should one immediately reject currently unorthodox ideas, such as linking the Proto-Finno-Ugric languages to the 'Periglacial refugium' period, as some have suggested (e.g. Wiik 2000; Dolukhanov 2000)?

Below we discuss a few 'case studies', in order to illustrate why does it seem to us that the genetic interpretation of demographic and probably linguistic histories needs a time span at least back to the late Pleistocene. From these case studies we wish, first of all, to illustrate what can be considered as one of the key questions in the understanding of present-day variation in the human mtDNA pool in Europe. Namely, bearing in mind a detailed topology of the mtDNA phylogenetic tree, what does it tell us about the beginnings of expansion of various of its sub-clusters? Following a phylogeographic approach, can we perhaps classify such signals over a variety of haplogroups and geographic regions? More to the point: are there significant differences in expansion times for mtDNA monophyletic clades, specific for different regions of Europe? And if yes, are there any patterns?

Out of a large number of possible examples, we first concentrate on U4 and U5 within a major western Eurasian haplogroup U. Both of them are recognized as Upper Palaeolithic, largely European varieties of mtDNA (Macaulay *et al.* 1999; Richards *et al.* 2000).

Case study: U5

U5 is a 'prototype' western Eurasian lineage cluster with a coalescence age of around 45,000-55,000 BP (Richards *et al.* 2000). Its phylogenetic tree does not suggest a star-like expansion from the founder (e.g. Richards *et al.* 2000; Finnila *et al.* 2001). But our analysis of a large number of U5 mtDNAs revealed the presence of about a dozen putative sub-founders, most of which exhibit nice star-like expansions (Fig. 35.1). More importantly, we have found that almost all of them exhibit coalescence ages around 11,000-13,000 BP and only a few, like 'the Saami U5', seem to have started to expand significantly more recently. The 'Saami motif — 16,270; 16,189, 16,144 — (labelled as S in Fig. 35.1, but the figure is here drawn without Saami variants) is rare outside northeastern Scandinavia: its topology in Scandinavia suggests recent severe bottleneck(s) in the demographic history of their carriers.

This nearly synchronous series of coalescence ages makes sense: it is much easier to imagine [specifically for a such an ancient branch) that an expansion phase hit all U5 twigs and limbs nearly simultaneously, than to assume a complicated pattern of a dozen or so widely irregular beginnings. Specifically, before the beginning of farming, a likely reason for an expansion may be traced to favourable climatic changes. A particular time frame before the beginning of the Holocene is highly likely, since this period corresponds (Younger Dryas and other cold events here excluded) to a rapid warming of climate after the LGM, and, therefore, to a re-occupation of large areas of northern Europe by humans as well varieties of animals and plants. We return to this question below.

Case study: U4

U4 is even more 'European' than U5: while U5 is relatively frequent all over western Eurasia, U4 is,

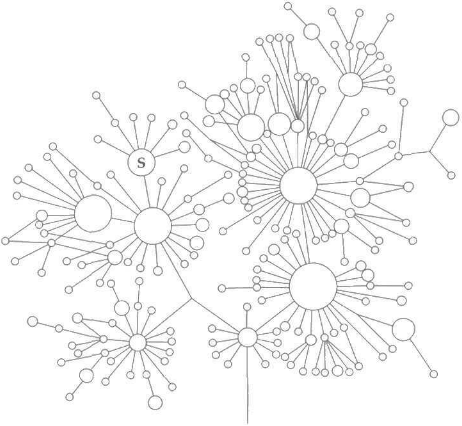

Figure 35.1. *Skeleton topology of the human mtDNA haphgroup U5 hypervariable 1 phylogenetic tree for western Eurasia. Circle sizes are proportional to numbers of individuals per haplotype. 'S' corresponds to a haplotype, most frequent among Saami population (HVR 1 motif 16,144; 16,189; 16,270).*

with a few interesting exceptions, more frequent in eastern Europe and is either absent or very rare in the Near East and elsewhere. In the European north, an interesting exception is the Saami mtDNA pool, where U4 is virtually absent.

We have constructed a HVRl-based phylogenetic tree for U4, using information from ~80 populations comprising a total of ~400 U4 genomes (Fig. 35.2). The topology of the U4 cluster is relatively simple, revealing the presence of a limited number of sub-founders. Of these, U4a and U4b are likely monophyletic, while U4c, determined by a transition at np 16,362, might be polyphyletic, at least in a pan-western Eurasian context. The highest frequencies of U4 (both in absolute terms and as a percentage of Hg U) can be observed actually not in Europe, but among Obi-Ugric Khantys and Mansis, living in northwestern Siberia. It is also frequent among the Finnic-speaking populations and in Volga Basin Turkic speakers, where, in some instances, its frequency exceeds that of U5,

In spite of this, we have not found any U4b mtDNA genomes among Finno-Ugric and Volga region people *(N* > 1000). This sub-cluster is largely, though not solely, typical for Germanic-speaking populations, being yet another highly characteristic example of a steep cline in the distribution of maternal lineages in Europe. The coalescence age of U4 is around 16,000-24,000 BP (Richards *et al.* 2000). With geographically more representative data at hand, it is interesting to estimate coalescence ages not only for each sub-division of U4 (i.e. U4*, U4a and U4b), but also for different linguistic/geographic entities within a sub-cluster. We found the answers intriguing. For the Baltic Finno-Ugric and Volga people (note that Hungarians differ here from the Finnic-speaking people), the coalescence ages both for U4* and U4a are around the maximum of the LGM, at 20,000-22,000 BP. Furthermore, taking U4c tentatively as monophyletic for this particular region, the corresponding sub-clade lineages in the FU-Volga area coalesce at about 19,000 BP — coinciding within the limits of error with U4* and U4a.

Postulating the beginning of expansion during the LGM seems strongly counter-intuitive at first glance. Here, however, comes an equally unexpected archaeological finding (Dolukhanov 2000), that actual population density (calculated from the number of precisely dated settlements) rose considerably after about 25,000 BP in a periglacial area of northern Ukraine-southern Russia, reaching its maximum around the peak of the LGM. It is of course highly speculative, but nevertheless tempting to bring these two completely independent findings together and to suggest that the eastern 'Periglacial refugium' postulated by archaeological data, and beginning of the expansion of U4 among eastern Europeans, can be attributed to the same prehistoric people.

Next, we calculated coalescence ages of U4*, U4a and U4b for the Germanic-speaking people (Germans, Norwegians, Swedes, Icelanders, Scots, German-speaking Swiss). Here the other interesting observation came. As indicated above, they share U4* and U4a with FU-Volga people, whereas the latter lack U4b. Nevertheless, the coalescence ages for all three indicated clades/sub-clades for the Germanic-speaking people are close and lie around 10,000-14,000 BP, suggesting that the beginning of their expansion was:

a) in the late Pleistocene, corresponding to the period of fast regression of continental ice cover in northern Europe and the general 'improvement' of climate;
b) much later (for U4* and U4a) than for people living in the adjacent geographic area of northeastern Europe.

One may ask about U4 in Mediterranean Europe. While U4a is so rare there that no meaningful calculation can be performed, the coalescence age for U4* for the Mediterranean is again about 13,000 BP.

Notice that this time scale (late Pleistocene) overlaps with that which one observes for the majority of the nicely star-like sub-clades of U5, discussed above. There is, of course, *a* profound difference between the spreads of U4 and U5 in western Eurasia. U5 is one of the major pan-western Eurasian maternal lineage clades, present in northwestern Africa, in the Near and Middle East and in Central Asia, while U4 is largely a northeastern-central European variety of mtDNA, found also in western Siberia/Altai and, in low frequencies, in Mediterranean Europe and the Near East. Interestingly, we have found a few U4 lineages even in India (see Kivisild *et al.* this volume).

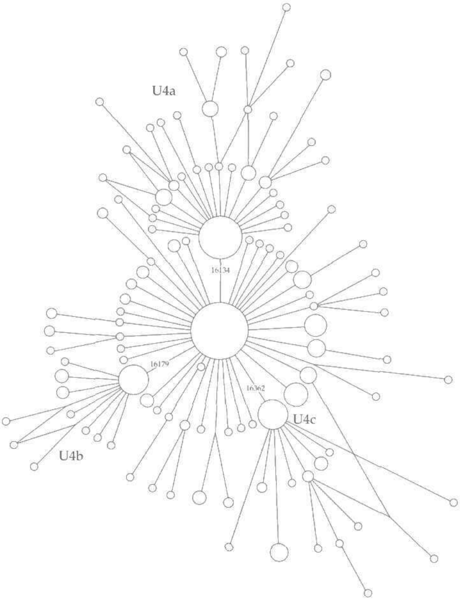

Figure 35.2. *Skeleton topology of the human mtDNA haplogroup U4 hypervariable 1 phylogenetic tree. The three sub-founders discussed (U4a, U4b and U4c) differ from the central node (referred as U* in the text) as indicated. About 400 U4 mtDNAs from about 15,000 mtDNAs were used to construct the tree.*

This pattern of a late Pleistocene expansion of population might be a general one. The coalescence age of largely European-specific (in contrast to western Eurasian U5) haplogroup V lies in late Pleistocene as well, and its expansion is explained in terms of the re-peopling of those parts of Europe deserted during the LGM (Torroni *et al.* 1998). For Hg V, this expansion seems to radiate from the Iberian peninsula (Torroni *et al.* 1998). Here, it is fair to refer to recent debate on this issue (Simoni *et al.* 2000; Torroni *et al.* 2000), clarified recently (Torroni *et al.* 2001). Furthermore, in Richards *et al.* (2000), coalescence ages of several other mtDNA haplogroups / sub-clusters display summary coalescence ages within a 'Mesolithic' time span (defined in this paper to lie approximately between 9000 and 14,000 BP).

To conclude; both for mtDNA and for the Y chromosome (e.g. Semino *et al.* 2000; Malaspina *et al.* 2000), it begins to appear that, although many lineage clusters currently visible in the European gene pool were likely present in Europe before the LGM, their present-day phylogeography is largely determined by demographic events after the LGM — during a period when re-peopling *before farming* seems to emerge as one of the most profound determinants of the present-day variability of European maternal lineages.

Several authors have rightfully stressed that a lineage cluster for individuals living in a given contiguous area cannot be assumed to have developed its diversity entirely or even largely in the area where it is found at present (e.g. Barbujani & Bertorelle 2001). In many instances it can be demonstrated that such cannot be so, in principle. After all, many northern European populations live at present in areas repopulated only after the LGM. This may mean that for Hg U5, widely spread also in the Near and Middle East, one can argue that its variability in Europe (which suggests coalescence ages for its sub-clades around 11,000-14,000 BP, see above) had already developed to a large extent in the Levant, Anatolia, the Caucasus, even in Egypt — wherever U5 is present today. All that is required is to assume that the founding migrants should have been numerous enough to carry the variants with them to Europe. Problems like this have been quantitatively treated in Richards *et al.* (2000), but it appears that the findings now seem to support the hypothesis that there was near-simultaneous expansion of sub-clusters of U5 around the late Pleistocene (in the late Upper Palaeolithic-Mesolithic), *together* with Hg V and sub-clusters of U4 among people whose maternal lineage descendants include Germanic language group Europeans. Because all three haplogroups are of a likely 'European' origin, the observation lends strong credence to a pan-European, late Pleistocene population expansion, traceable to the re-peopling of the region after the LGM.

Quite a few more examples may be added. Within European haplogroup H, we have:

1. a sub-cluster characterized by an additional transition at np 16,261 with a star-like phylogenetic tree. From its pan-European data bank (this lineage cluster is very rare outside Europe) we have at present, with $N = 79$, a coalescence age of 12,900±1800 BP; for the FU-Volga area 12,300±3300 BP, and for Germanic language group populations 12,800±3300 BP;
2. a sub-cluster of Hg H, characterized by an additional transition at np 16,209 with a coalescence age of 12,600±3000 BP.

A detailed knowledge of the phylogeography of U4 offers several other avenues for speculation. It is nearly absent among Iranians, Ossetes and Kurds — i.e. among three Indo-Iranian-speaking populations. Neither have we found it in a limited sample of Tadjiks, yet another Indo-Iranian-speaking population. In contrast, U4 is present among some South Caucasus populations, in particular among Kartvels (Georgians), but again infrequent among Turks (Tambets *et al.* 2000). Although there is not enough space to discuss all of the conclusions stemming from these findings here, they seem to suggest the absence of any massive maternal gene flow from eastern Europe (where U4 is one of the oldest and most frequent mtDNA clades) to the Middle East since the end of the LGM. One may add that the opposite is also doubtful, argued on the basis of a near, although not complete, absence of U7 in northern and eastern Europe, while this cluster is most frequent worldwide in the Middle and Near East as well as in western India (Kivisild *et al.* 1999; see also Kivisild *et al.* this volume).

Nevertheless, if one favours scenarios with massive migrations from the Middle East to eastern Europe (present in many maps as a bold arrow originating in the Middle East, turning around the eastern Caspian and pointing west), then it is necessary to postulate that U7 reached Iran only after this putative northwards expansion. This is not an easy and straightforward assumption, because, as already mentioned, Iran is rich in U7, at least at present, and there is no obvious 'homeland' for U7 elsewhere. On the other hand, the presence of U7 in the Balkans suggests gene flow into Europe via Anatolia.

Some 'mini-cases' signal expansions in the Neolithic and Bronze Age

From the previous case studies, one may obtain an impression that we wish to synchronize the coalescence ages of all mtDNA lineage clusters/sub-clusters/sub-sub-clusters with demographic events in the Pleistocene, in particular with the post-LGM re-peopling of Europe. That is not so. One complication with more recent events is that they might well be 'hidden' — i.e. assimilated as 'new boosts' with the beginning of food production — into expansions which began earlier, in the late Pleistocene. Therefore, more recent events (expansions) can be best detected in cases where *de novo* sub-founders arose and produced star-like expansions.

With some fear of making the issue even more complex, we add here several such 'mini-cases' exhibiting coalescence ages well within the time span of food production.

Sub-haplogroup Tl - HVR I motif from T -16,163; 16,186; 16,189*

1. For the European Finno-Ugric and Volga-South Ural Turkic-speaking people, the coalescence age of sub-haplogroup Tl is about 4000 BP. However, it appears to be even more recent for the Baltic-Finnic people (Finns, Estonians, Karelians). If these are removed from the calculations, then the coalescence age of Tl for the Volga-Uralic people (here: Maris, Mordvin, Komis, Udmurts, Tatars, Chuvashis, Bashkirs) rises to ~6500 BP.
2. For Germanic-language group people plus Scots, the coalescence age of sub-haplogroup Tl is about 4000 BP. This calculation ignores all putative reversions in mtDNA HVR 1 sequence data bases, If one accepts them, the coalescence age rises to about 6500 BP.
3. For Anatolia, South Caucasus, the Near and Middle East and northeast Africa, the overall coalescence age for Tl is very much older — around 17,000 BP.

Sub-haplogroup Tl is one of the clearest examples of a lineage cluster with a much earlier expansion in the Near and Middle East and South Caucasus than in Europe. Most importantly, it may possibly testify to the arrival of Neolithic farmers in Europe (e.g. Metspalu *et al.* 1999; Tambets *et al.* 2000; Richards *et al.* 2000). However, our new inspection, based on a more extensive study of the Near Eastern mtDNA pool, shows that the 'really old' part for Tl within western Asian seems to be the South Caucasus, Anatolia and Syria, possibly including also southeastern Mediterranean Europe and Egypt, where the corresponding tree coalesces about 26,000-29,000 BP. Furthermore, like the late Pleistocene series of coalescence ages for U5 and U4 in Europe, the western Asian coalescence ages for Tl pre-date significantly the beginnings of farming and animal domestication in Anatolia and the Fertile Crescent and may have been trigged by an early post-LGM climatic change, or by even more remote and unknown events before the LGM.

Interpretation of the coalescence ages for Tl in regionally/linguistically divided Europe is also not as obvious. While there seems to be one time frame corresponding roughly to an early expansion of agriculture, there are also signs of a significantly later expansion in northern Europe. There is not yet a powerful-enough data base for northern Russians, but the coalescence age for Tl among northeastern Komis and Udmurts is as recent as it is for the Baltic Finnic-speaking people — about 3000 to 4000 BP. This suggests a late arrival of females carrying Tl mtDNA to northeastern Europe.

Sub-duster of Hg J - 16,069; 16,126; 16,145; 16,261; 16231

Although haplogroup J belongs to the list of putative Neolithic arrivals, this particular sub-clade of J is yet another largely 'northern European' variety of mtDNA lineage, very probably arisen in Europe and its coalescence age is around 7000 BP (N = 58; 7000±1600 BP). This might be an overestimate: subtracting a putative sub-clade node at np 16,189, the coalescence age drops to about 5000 BP.

Thus, provided a sufficiently large data bank is at hand, one may 'collect' a long list of 'mini-clades' which coalesce around 4000-7000 BP. A relatively large standard deviation makes any detailed interpretation of the results complicated, but taking 4000-6000 BP as an average, the coalescence ages (signs of the beginning of an expansion of a particular clade) do fall into a time frame, corresponding to the 'consolidation and further expansion' of agriculture in Europe (Pinhasi *et al.* 2000). Identifying such mini-clades all over the main mtDNA lineage clusters supports a suggestion that we are dealing here with a general phenomenon.

Concluding remarks

It might sound unimaginative, but it appears that for further significant progress in finding answers to the 'big question' (defined as by Renfrew 2000), it will be necessary to gather more data and better data. As far as genetics is concerned, the *best* data are those which allow a detailed phylogeographic analysis to be carried out. In other words, this means making the best use of a fortunate situation for the researcher who is dealing with non-recombining, uniparentally inherited DNA. At least at the present, the barrier to *knowability* in hopefully still some unknown distance ahead. One may expect that most remaining problems in the phylogenetic analysis of mtDNA data will soon reach maturation as far as the topology of the corresponding tree (at least for western Eurasia) is concerned, and that the problems identified here will be largely solved, or shown to remain ambiguous forever because of built-in limitations resulting from the length of mtDNA.

What is much less clear is how we can reach significantly better temporal resolutions. Take, for example, U5: a cluster coalescing around 40,000-50,000 BP but consisting, as we interpret it now, of a number of sub-founders coalescing about 12,000 BP. Even though U5 mtDNAs are frequent in the western Eurasian mtDNA pool, to identify numerous sub-clusters within it one does need to operate with large sample sizes. For less frequent mtDNA varieties, only very large data bases, consisting of data about tens of thousands of mtDNAs, will allow a detailed temporal analysis, in particular for a time frame when important farming/language dispersal events took place. But there is clearly a light visible and, as already mentioned (Torroni *et al.* 2000; Helgason *et al.* 2001), deeper phylogenetic analysis reveals that there is no uniformity in the spread of maternal lineages in Europe. Now that genetics is approaching an increasingly finer phylogeographic resolution, including the temporal component, it depends increasingly on better coverage and resolution of time and density maps of archaeological sites (e.g. Pinhasi *et al.* 2000; Zvelebil this volume).

References

Bandelt, H.-J., P. Forster, B.C. Sykes & M.B. Richards, 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141, 743-53.

Barbujani, G. & G. Bertorelle, 2001. Genetics and the population history of Europe. *Proceedings of the National Academy of Sciences of the USA* 98, 22-5.

Bellwood, P., 2001. Early agricultural diasporas? Farming, languages and genes. *Annual Review in Anthropology* 30, 181-207.

Cann, R.L., M. Stoneking & A.C Wilson, 1987. Mitochondrial DNA and human evolution. *Nature* 325, 31-6.

Dolukhanov, P., 2000. 'Prehistoric revolutions' and languages in Europe, in *The Roots of Peoples and Languages of Northern Eurasia II and III,* ed. A. Künnap, Tartu: Tartu University Press, 71-84.

Finnilä, S., M.S. Lehtonen & K. Majamaa, 2001. Phylogenetic network for European mtDNA. *American Journal of Human Genetics* 68, 1475-84.

Forster, P., R. Harding, A. Torroni & H.-J. Bandelt, 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *American Journal of Human Genetics* 59(4), 935-45.

Helgason, A., E. Mickey, S. Goodacre, V. Bosnes, K. Stefansson, R. Ward & B. Sykes, 2001. MtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *American Journal of Human Genetics* 68, 723-37.

Jacob, F., 1982. *The Possible and the Actual.* New York (NY): Pantheon Books.

Kivisild, T., M.J. Bamshad, K. Kaldma, M. Metspalu, E, Metspalu, M. Reidla, S. Laos, J. Parik, W.S. Warkins, M.E. Dixon, S.S. Papiha, S.S. Mastana, M.R. Mir, V. Ferak & R. Villems, 1999. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Current Biology* 9,1331-4.

Macaulay, V.A., M.B. Richards, E. Hickey, E. Vega, F. Cruciani, V. Guida, R. Scozzari, B. Bonné-Tamir, B. Sykes & A. Torroni, 1999. The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *American Journial of Human Genetics* 64(1), 232-49.

Malaspina, P., F. Cruciani, P. Santolamazza, A. Torroni, A. Pangrazio, N. Akar, V. Bakalli, R. Brdicka, J. Jaruzelska, A. Kozlov, B. Malyarchuk, S.Q. Mehdi, E. Michalodimitrakis, L. Varesi, M.M. Memmi, G. Vona, R. Villems, J. Parik, V. Romano, M. Stefan, M. Stenico, L. Terrenato, A. Novelletto & R. Scozzari, 2000. Patterns of male-specific inter-population divergence in Europe, West Asia and North Africa. *Annals of Human Genetics* 64(5), 395-412.

Metspalu, E., T. Kivisild, K. Kaldma, J. Parik, M. Reidla, K. Tambets & R. Villems, 1999. The Trans-Caucasus and the expansion of the Caucasoid-specific human mitochondrial DNA, in *Genomic Diversity,* eds. S. Papiha, R. Deka & R. Chakraborty. New York (NY); Kluwer Academic/Plenum Publishers, 121-34.

Pinhasi, R., R.A. Foley & M.M. Lahr, 2000. Spatial and temporal patterns in the Mesolithic-Neolithic archaeological record of Europe, in Renfrew & Boyle (eds.), 45-56.

Renfrew, C, 2000. At the edge of knowability: towards a prehistory of languages. *Cambridge Archaeological Journal 10(1),* 7-34.

Renfrew, C. & K. Boyle (eds.), 2000. *Archaeogenetics: DNA and Population Prehistory of Europe.* (McDonald Institute Monographs.) Cambridge: McDonald Institute for Archaeological Research.

Richards, M. & V. Macaulay, 2001. The mitochondrial gene tree comes of age. *American Journal of Human Genetics* 68, 1475-84.

Richards, M., H. Corte-Real, P. Forster, V. Macaulay, H. Wilkinson-Herbots, A. Demaine, S. Papiha, R. Hedges, H.-J. Bandelt & B. Sykes, 1996. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *American Journal of Human Genetics* 59(1), 185-203.

Richards, M., V. Macaulay, E. Hickey, E. Vega, B. Sykes, V. Guida, C. Rengo, D. Sellitto, F. Cruciani, T. Kivisild, R. Villems, M. Thomas, S. Rychkov, O. Rychkov, Y. Rychkov, M. Golge, D. Dimitrov, E. Hill, D. Bradley, V. Romano, F. Calf, C. Vona, A. Demaine, S. Papiha, C. Triantaphyllidis, G. Stefanescu, J. Hatina, M. Belledi, A. di Rienzo, A. Novelletto, A. Oppenheim, S. Norby, N. Al-Zaheri,

S. Santachiara-Benerecetti, R. Scozzari, A. Torroni & H.-J. Bandelt, 2000. Tracing European founder lineages in the near eastern mtDNA pool. *American Journal of Human Genetics* 67,1251-76.

Saillard, J, P. Forster, N. Lynnerup, H.-J. Bandelt & S. Norby, 2000. MtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *American Journal of Human Genetics* 67(3), 718-26.

Semino, O., G. Passarino, P.J. Oefner, A.A. Lin, S. Arbuzova L.E. Beckman, G. de Benedictis, P. Francalacci, A. Kouvatsi, S. Limborska, M. Marcikiae, A. Mika, B, Mika, D. Primorac, A.S. Santachiara-Benerecetti, L.L. Cavalli-Sforza & P.A. Underhill, 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y-chromosome perspective. *Science* 290, 1155-9.

Simoni, L., F. Calafell, D. Pettener, J. Bertranpetit & G, Barbujani, 2000. Geographic patterns of mtDNA diversity in Europe. *American Journal of Human Genetics 66, 262-78*.

Tarnbets, K., 2000. The topology of the maternal lineages of the Anatolian and Trans-Caucasus populations and the peopling of Europe: some preliminary considerations, in Renfrew & Boyle (eds.), 219-35.

Tambets, K., T. Kivisild, E. Metspalu, J. Parik, K. Kaldma, S. Laos, H.-V. Tolk, M. Golge, H. Demirtas, T. Geberhiwot, S. Papiha, S., G.F. De Stefano & R. Villems, 2000. The topology of the maternal lineages of the Anatolian and Trans-Caucasus populations and the peopling of the Europe: some preliminary considerations, in Renfrew & Boyle (eds.), 219-35.

Torroni, A., H.-J. Bandelt, L. D'Urbano, P. Lahermo, P. Moral, D. Sellitto, C. Rengo, P. Forster, M.L. Savontaus, B. Bonné-Tamir & R. Scozzari, 1998. MtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *American Journal of Human Genetics* 62(5), 1137-52.

Torroni, A., M. Richards, V. Macaulay, P. Forster, R. Villems, S. Nørby, M.L. Savontaus, K. Huoponen, R. Scozzari & H.-J. Bandelt, 2000. MtDNA haplogroups and frequency patterns in Europe. *American Journal Of Human Genetics* 66(3), 1173-7.

Torroni, A., H.-J. Bandelt, V. Macaulay, M. Richards, F. Cruciani, C. Rengo, V. Martinez-Cabrera, R. Villems, T. Kivisild, E. Metspalu, J. Parik, H.V. Tolk, K. Tambets, P. Forster, B. Karger, P. Francalacci, P. Rudan, B. Janicijevic, O. Rickards, M.L. Savontaus, K. Huoponen, V. Laitinen, S. Koivumaki, B. Sykes, O. Hickey, A. Novelletto, P. Moral, D. Sellitto, A. Coppa, N. Al-Zaheri, A.S. Santachiara-Benerecetti, O. Semino & R. Scozzari, 2001. A signal, from human mtDNA, of postglacial recolonization in Europe. *American Journal of Human Genetics* 69(4), 844-52.

Valladas, H., J. Clottes, J.-M. Geneste, M.A. Garcia, M. Arnold, H. Cachier & N. Tisnerat-Laborde, 2001. Palaeolithic paintings: evolution of prehistoric cave art. *Nature* 413,479.

van Andel, T.H., 2000. Where received wisdom fails: the mid-Palaeolithic and early Neolithic climates, in Renfrew & Boyle (eds.), 31-9.

Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes & A.C. Wilson, 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253, 1503-7.

Wallace, D., 1995. Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *American Journal of Human Genetics* 57,201-23.

Wiik, K., 2000. Some ancient and modern linguistic processes in northern Europe, in *Time Depth in Historical Linguistics*, eds. C. Renfrew, A. McMahon & *L.* Trask. (Papers in the Prehistory of Languages.) Cambridge: McDonald Institute for Archaeological Research, 463-79.

Zvelebil, M., 2000. The societal context of the agricultural transition in Europe, in Renfrew & Boyle (eds.), 31-9.

MCDONALD INSTITUTE MONOGRAPHS

Examining the farming/ language dispersal hypothesis

Edited by Peter Bellwood & Colin Renfrew