Richard Villems, Siiri Rootsi, Kristiina Tarnbets, Maarja Adojaan, Vladimir Orekhov, Elsa Khusnutdinova, Nikolay Yankovsky

Archaeogenetics of Finno-Ugric speaking populations

Summary

In our third paper of this series (see Villems et al. 1998; Rootsi et al. 2000) we continue our quest for the genetic origin of Finno-Ugric speaking people and extend our previous analysis of patrilinearly (via Y chromosome) and matrilinearly (via mitochondrial DNA) inherited genes in the Eurasian gene pool of humans. We discuss new data of others as well as ourselves about the westem Eurasian populations and, in particular, those living in Volga basin and the southern Ural region. The conclusions we reach strengthen further previous ideas about a common Palaeolithic mtDNA pool of Europeans, extending to the Ural Mountains. A sharp cline becomes apparent in the southern Ural – western Kazakhstan area, suggesting a predominantly western Eurasian substratum in the European Turkic-speaking populations such as Volga Tatars and Chuvashis, whereas maternal lineages of their immediate neighbors – Bashkirs and Kazakhs – are already largely of an eastern Eurasian origin. We also portrait here the frequency pattern of the Y-chromosomal lineage clusters of 4 Finno-Ugric populations and of 7 from neighboring areas, to reveal both common as well as specific features in the paternal genetic background of the characterized populations.

Introduction

Since the "African Eve" (Cann et al. 1987), the usefulness of the analysis of the variability of the matrilinearly inherited mitochondrial genome of the extant human populations to understand "from where we are" became apparent and the last five years in human population genetics is a clear demonstration of the heuristic power of the investigation of two, strictly speaking single, genetic *loci:* of the mitochondrial genome (mtDNA) and of the Y chromosome. Fine resolution of the human maternal lineages started to emerge already earlier (first reviewed in Wallace 1995), but the global view to our patrilinear phylogeny is now catching up rapidly (Underhill et al. 2000). The main outcome of the analysis of both genetic systems is encouragingly equivocal: modem humans arose in Africa and started to colonize the other continents according to the "recent out-of-Africa" scenario. This genetic evidence based conclusion had a profound influence to our understanding of the descent of human beings.

More locally – for Europe and for western Eurasia in general, a set of different conflicting views on the peopling of Europe by modem humans were hold until very recently. The first and most profound was the problem of the placing of Neanderthals – did this extinct variety of *Homo sapience* contribute to our genetic pool? It seems now that the answer preferred by geneticists is "no" (for the discussion of morphological evidence see Niskanen 2000). First, it was shown that the Euro-

pean mtDNA gene pool does not contain lineages, possibly descending to us from Neanderthals (Torroni et al. 1994; Richards et al. 1996). Analysed by now fragments of mtDNA of three different Neandertal fossils confirmed convincingly this prediction (Krings et al. 1997; Ovchinnikov et al. 2000; Krings et al. 2000). It is fair to admit that this negative evidence comes from a single genetic locus – mtDNA; successful amplification of relevant portions of Y-chromosomal DNA would be most useful (but extremely problematic to be achieved); not to add that possible Neanderthal mtDNA lineages may have been lost by drift. As far as autosomal genes are concerned, the issue might be even more complex to explore: thanks to 3 4 fold higher effective population size for the latter, the coalescence age for many such genes may well pre-date the most recent common ancestor (MRCA) of modem humans and Neanderthals. All in all, the important fact is that the genetic evidence for admixture is absent and it is certainly reasonable to postpone further discussions until such evidence (or an equivalent morphological proof) is found, if ever.

The second and equally extensively debated problem is the identity of the extant Europeans in a sense whether the contemporary population is formed largely by descendants of recent (Neolithic) migration(s) from the Near East – Gordon Childe's ***Ex Oriente*** Lux – or are they rather the late Upper Palaeolithic modem humans, who re-expanded in Europe after the Last Glacial Maximum? Because the history of this debate was well covered in previous issues of this series (e.g. Dolukhanov 2000; Wiik 2000; Niskanen 2000; Villems et al. 1998; Rootsi et al. 2000), we briefly mention here only its essence and refer to some recent papers on this topic. As it is only too well known, principal component analysis of the distribution of classical genetic markers revealed a wave-like pattern of their frequencies (summarised in Cavalli-Sforza et al. 1994). In particular, the first principal component suggested a gradient of the distribution of gene frequencies from Anatolia – Near East to western Europe. Although such a wave lacks direction, it overlapped with a pattern of archaeological evidence, suggesting a slow advance of agriculture from the Levant. Bringing these data together, a Neolithic demic diffusion hypothesis was postulated, including concurrent spread of Indo-European languages into Europe. In 1996, the first large study covering European maternal lineages was published (Richards et al.). In this paper, a radically different conclusion was reached: an overwhelming majority of the found mtDNA lineages are in Europe since Palaeolithic. The main argument in favour of this conclusion was the found deep coalescence time for most of the European mtDNA lineage clusters (haplogroups), pre-dating the beginning of Neolithic and, consequently, of agriculture in Europe. Although the mtDNA is just a single genetic locus, it nevertheless reflects the genetic history of a half of population; consequently – it was justified to reach a general conclusion that the contemporary Europeans are largely descendants of the people who have lived here in the Palaeolithic epoch and probably throughout the Last Glacial Maximum (LGM), about 22,000 – 18,000 years before present (BP). This main conclusion was strongly criticised soon (Cavalli-Sforza and Minch 1997).

More recently, a series of new papers appeared, questioning a predominant "Palaeolithicity" of contemporary Europeans. Using nuclear DNA markers and applying a spatial autocorrelation method, Chikhi and colleagues (1998a and 1998b) reached a radically opposite conclusion, according to which only a very minor part, if any, of the contemporary European nuclear gene pool is from Palaeolithic or even Mesolithic and that the people carrying the existing variety we observe at present, entered Europe during the Neolithisation, likely from the Near East. Furthermore, from their re-analysis of the large sample of mtDNA HVR 1 sequence clusters (Simoni et al. 2000) drew a conclusion that the previous (Torroni et al. 1998) interpretation of the Mesolithic re-population of Europe from refugia, present in a limited places in western and eastern Europe (specifically concerning haplogroup V and the Franco-Cantabrian refugium) after the LGM, is unfounded. The

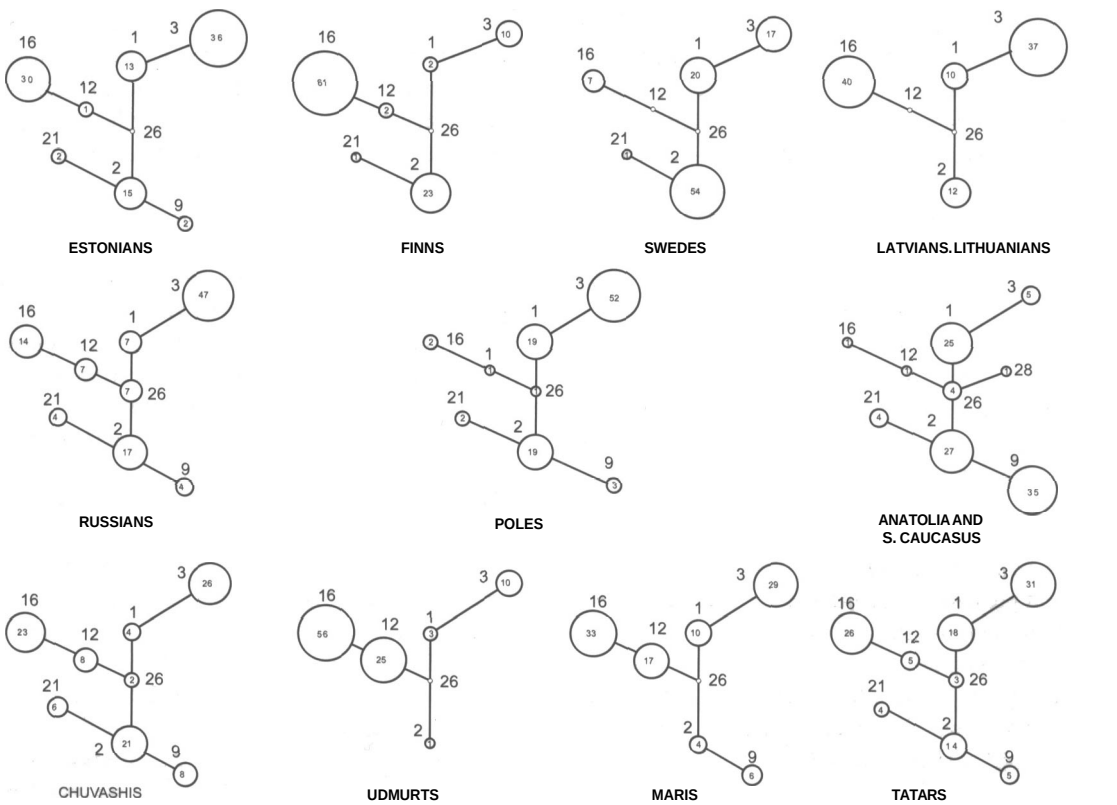

Figure 1. Frequencies (%) ***of Y-chromosomal haplogroups of several western Eurasian populations. Estonians, Chuvashis, Udmurts, Tatars, Poles, Caucasus and Anatolia region – studied by us; Russians, Maris, Finns, Latvians and Lithuanians (Rosser et al. 2000); Swedes (Helgason et al. 2000). Numbers a) outside circles: haplogroups; b) inside circles: per cent from total.***

latter paper received strong criticism (Torroni et al. 2000), but the basic problem about the "historical identity" of the genes of the contemporary Europeans remains.

One may nevertheless be optimistic. The strength of using unipaternally inherited genetic loci (mtDNA and Y chromosome) lies in the possibility to restore their "true geneology". The resulting phylogenetic trees, in particular a detailed enough understanding of their topology, will hopefully allow to reach a consensus view and this day might not be even far away. There are several aspects in this analysis, worthwhile to mention. First of all, one needs to consider more complex scenarios, involving back migrations from Europe to Near and Middle East. Historically, such migrations are known at least in some occasions (e.g. in late Iron Age) and one can assume that hectic temperature shifts during the Upper Palaeolithic (and later), both before and after the LGM, resulted in several migration waves back to more milder climate as well. A systematic attempt to take such back and forth migrations into account was recently proposed, together with a search for "founder lineages" (Richards et al. 2000).

The other reason for optimism is in a recent rapid progress in the understanding of the European (Rosser et al. 2000; Semino et al. 2000) and of the world (Underhill et al. 2000) phylogeny of Y chromosome. Most importantly, the new Europe-

an Y-chromosomal data is interpreted in a way, very close to that for mtDNA: both lines of evidence suggest that around 80 per cent of human genetic lineages in Europe are here already since Palaeolithic. Furthermore, the new Y-chromosomal data suggests a post-LGM repopulation/relocation inside Europe (Semino et al. 2000), in conformity with similar suggestions about the most of mtDNA lineages (Torroni et al. 1998; Richards et al. 2000).

Genetic legacy of the eastern European populations

In order to reconstruct a genetic history of European populations, probably the most urgent problem at present is to bring the knowledge about the central, south-eastern and eastern part of Europe to the level, currently available for the western part of the Continent (although even there one can see numerous smaller gaps). Unless that is achieved, answers to many, if not to all more general questions remain pending. In a specific "Northern" context, this is absolutely necessary for the understanding of the genetic history of people, speaking Finno-Ugric languages. One may add, however, that equally poorly are understood populations, providing an essential context for Finno-Ugrians – first of all eastern Slavic and Baltic populations and eastern European Turkic-speaking populations. There are some articles concerning the Baltic populations (Sajantila et al. 1995, 1996; Kittles et al. 1998; Villems et al. 1998; Lahermo et al. 1999; Rootsi et al. 2000), a preliminary article about Russians (Orekhov et al. 1999) and also a limited dataset for Volga region Finno-Ugric speakers (Sajantila et al. 1995, Lahermo et al. 1999). Recent European Y chromosomal studies (Malaspina et al. in press; Rosser et al. 2000; Semino et al. 2000), have included some populations from eastern parts of Europe. All that, taken together, is a good starting position for a more comprehensive research ahead.

Together with our Russian colleagues we have now initiated a more systematic look to an area near to west and south of the Ural Mountains: to the genes of Finno-Ugric speaking populations from Volga-Finnic and Permian linguistic branches and to their close geographical neighbors, who linguistically belong to the Turkic branch of Altaic languages. This work is currently in progress and below we draw some preliminary conclusions, which serve also as working hypotheses for further studies. One may add that some other groups are currently concentrating to Siberian populations, including Obi-Ugric and Samoyedic language group people. The results to be obtained will certainly facilitate the reconstruction of the genetic history of Finno-Ugric speaking people considerably.

Phylogeography of the European patrilinear inheritance

In the previous issue of this series (Rootsi et al. 2000), we mostly concentrated on phylogeography of one specific variant of Y chromosomes, called *Tat **C*** allele or haplogroup 16 in a nomenclature of C. Tyler-Smith and M. Jobling. In this article we pay the main attention to the populations of Volga Basin, comparing them with several relevant western Eurasian populations studied by us and others. In addition, the haplogroup distribution data of Maris, published in Rosser et al. (2000), is also included to the present analysis.

A Y-chromosomal *haplogroup* (hg) is determined by combinations of single nucleotide polymorphism~and is placed into a tree in a cladistic way. Different lineages inside a haplogroup, which are described by length variants of studied STRs, are named *haplotypes*. The number of different haplotypes inside haplogroups, as well as their frequencies, enables to study and compare genetic diversity of paternal lineages of different populations.

One difficulty that somewhat complicates the comparison of Y-chromosomal data sets from dif-

ferent sources, is the lack of common nomenclature so far. As already mentioned, several extensive studies about European Y chromosomes were published recently (Malaspina et al. in press; Rosser et al. 2000; Semino et al. 2000), in which haplogroups have been defined differently. The systems overlap only partially and attempts to bring them together result sometimes only in approximate solutions because different markers have been used. To distinguish variants of Y chromosomes as certain haplogroups we use here the nomenclature of Jobling et al. (1997).

Haplogroups 1 and 3 (correspond to haplotypes Eu18 and Eu19 in Semino et al. 2000) together cover more than 95 % of European Y chromosomes. Although phylogenetically closely related (hg 1 is an ancestral state of hg 3), geographical distribution of their frequencies forms opposite gradients. Haplogroup 1 dominates in western Europe, being frequent in Spain (68 %) and reaching maximum value in Ireland (81 %). Its frequency decreases towards east: for example, in Ukraine it is as low as 4 per cent (Rosser et al. 2000). Semino et al. (2000) suggested that the reason for such opposite clines for hg 1 and hg 3 is in an expansion from different population refugia after the LGM, i.e. during the post-LGM re-population of Europe. Accordingly, the expansion of hg 1 started from Iberia and that for hg 3, from the territory of the present-day Ukraine. Because hg 1 is relatively infrequent in populations of our direct interest here, we do not discuss its phylogeography further.

The Y chromosome profiles or haplogroup frequencies for some Eurasian populations are shown in Figure 1. The most frequent haplogroup among studied eastern-European populations is hg 3, in particular in Slavs (Russians and Poles), but also in Baltic populations (Indo-European speaking Latvians and Lithuanians as well as Finno-Ugric speaking Estonians). The same haplogroup is represented frequently also in all Volga basin populations: among Tatars, Chuvashis and Maris, i.e. irrespective of linguistic affinity of its carriers. At the same time, hg 3 frequency decreases in the direction of Finns and Swedes. The same occurs in the southern direction, towards Anatolia and the Caucasus. These results agree with data by Rosser et al. (2000), where a clear north-east/south-west cline is described. In the latter article, the spread of this haplogroup is considered to mark the movement of the Kurgan people from north of the Caspian Sea, dated around 7,000 YBP. In a sharp contrast, Semino et al. (2000) suggest that hg 1 and hg 3 have entered Europe already 40,000 – 35,000 years ago together with the Aurignacian culture. These differences in timing are very substantial and reveal perhaps the main remaining weakness of the current state of art in using the Y chromosome as a tool. However, there is still a huge potential in the refining of the topology of the phylogenetic tree and methods to calculate the speed of molecular evolution for the Y chromosome.

The clinal geographical distribution of hg 3, as already said, probably reflects the re-population of Europe after the LGM from a refugium in the territory of the present-day Ukraine (Semino et al. 2000). Rosser et al. (2000) suggest that the origin of hg 3 might be in the Ukraine. At the same time, haplogroup 3 is also frequent in Central Asia (Zerjal et al. 1999) and Pakistan (Underhill et al. 2000). Yet it is quite complicated to see how the northern Pontic Y-chromosomal pool could have played such a profound role in the establishing the extant Indian paternal lineages. Our unpublished analysis shows that the diversity of hg 3 is high also in Indian populations. The situation here has an analogue in mtDNA hg U, where an early Upper Paleolithic split divided this dominant western/southern Eurasian lineage cluster into two nearly non-overlapping parts between the western Eurasian and Indian populations (Kivisild et al. 1999). A higher resolution study of the "inner topology" of hg 3 is needed to resolve the problem – the existing tools (STRs) might not allow to reach an unambiguous answer because of their too short time horizon.

It is interesting to compare the Y-chromosomal diversity of Turkic speakers from Volga region and from Anatolia. While hg 9 dominates in Anatolia, then among Volga Tatars and Chuvashis this

hg comprises, like in other East-European populations, less than 10 per cent of the total (Fig. 1). Once again, it suggests that geographic neighborhood is often more important than linguistic relatedness. Even in situations where one can expect mainly male-dominating migrations – like the conquering of Anatolia by Turkish tribes less than a thousand years ago.

Haplogroups 16 and 12 in Volga basin: distribution and diversity

Haplogroup 16 or *Tat* C allele is practically missing in western European populations (Zerjal et al. 1997), but its frequency is high among Baltic-Finnic populations – in Estonians and Finns, and also in Finno-Ugric populations in Volga, such as Udmurts and Maris (Fig. 1). But that is not all: hg 16 is seen at medium to high frequencies in other neighboring populations as well: among Chuvashis and Tatars in Volga region and among Lithuanians and Latvians in the Baltic. The sharp cline in hg 16 frequencies between Poles and Lithuanians supports an earlier suggestion that the settlement area of proto-Finno-Ugric populations was once much wider (e.g. Indreko 1948). As we suggested already earlier (Villems et al. 1998; Rootsi et al. 2000), the combined higher frequency and diversity of hg 16 among populations inhabiting the eastern Baltic and Eastern European Plain indicates from where to find the "ancient homeland" of the *Tat* C mutation. We remind here that the nearly circum-Arctic spread of hg 16 has been discussed in several papers (Zerjal et al. 1997; Villems et al. 1998; Lahermo et al. 1999; Karafet et al. 1999; Rootsi et al. 2000).

Hg 12 is ancestral to hg 16 and its distribution overlaps to some extent with that for hg 16 (Zerjal et al. 1997; Rosser et al. 2000). Y chromosomes in this haplogroup are carrying a 50f12 deletion and differ from hg 16 chromosomes in having *Tat* T allele instead of *Tat* C. The highest frequency of hg 12 was detected in Volga Finno-Ugric populations – Udmurts and Maris (Fig. 1); Udmurts have the highest frequency of hg 12 among the so far investigated populations. Rosser et al. (2000) have suggested that the population of origin of the *Tat* mutation could be Maris, because according to their data hg 12 is most frequent in this population. However, because the frequency of hg 12 is even higher among Udmurts, one should leave the question open. Perhaps such an identification, applied to the extant populations, is in fact not appropriate anyway for a reason that, very likely, the particular mutations pre-date times since when one can speak about distinct sub-populations of the eastern European Finno-Ugric people.

Diversity of an SNP-established haplogroup can be studied by finding such additional SNPs within a haplogroup that allow to split a haplogroup into sub-groups, i.e. to divide a branch of the tree into limbs and twigs. The same can be achieved using any other markers, exhibiting diversity within a haplogroup. We have used 5 STRs (DYS 19, DYS 390, DYS 391, DYS 392 and DYS 393) and employed median network approach (Bandelt et al. 1995) in order to reconstruct the most parsimonious phylogenetic trees (Fig. 2). The present more extensive study confirms our earlier finding (Villems et al. 1998; Rootsi et al. 2000) that the STR length diversity in hg 16 carriers is higher in eastern Europeans than in the two studied Siberian populations (Yakuts and Buryats), where its variability is quite low (Zerjal et al. 1997). The diversity is equally high in Chuvashis and Tatars (Fig. 2).

For Udmurts, although they have the highest frequency of hg 16 as well as for hg 12, the variability within the two haplogroups is significantly lower, specifically so for hg 12 (Fig. 2). For example, whereas all four Chuvashi hg 12 Y chromosomes differ from each other, then 24 out of 25 Udmurt Y chromosomes belong to a single haplotype. In this respect, the pattern of the variability of hg 16 in Udmurts is similar to that of hg 16 for Buryats and Yakuts, discussed by us in detail earlier (Rootsi et al. 2000). This probably suggest an influence of a very pronounced genetic drift – a recent massive founder effect orland bottleneck –

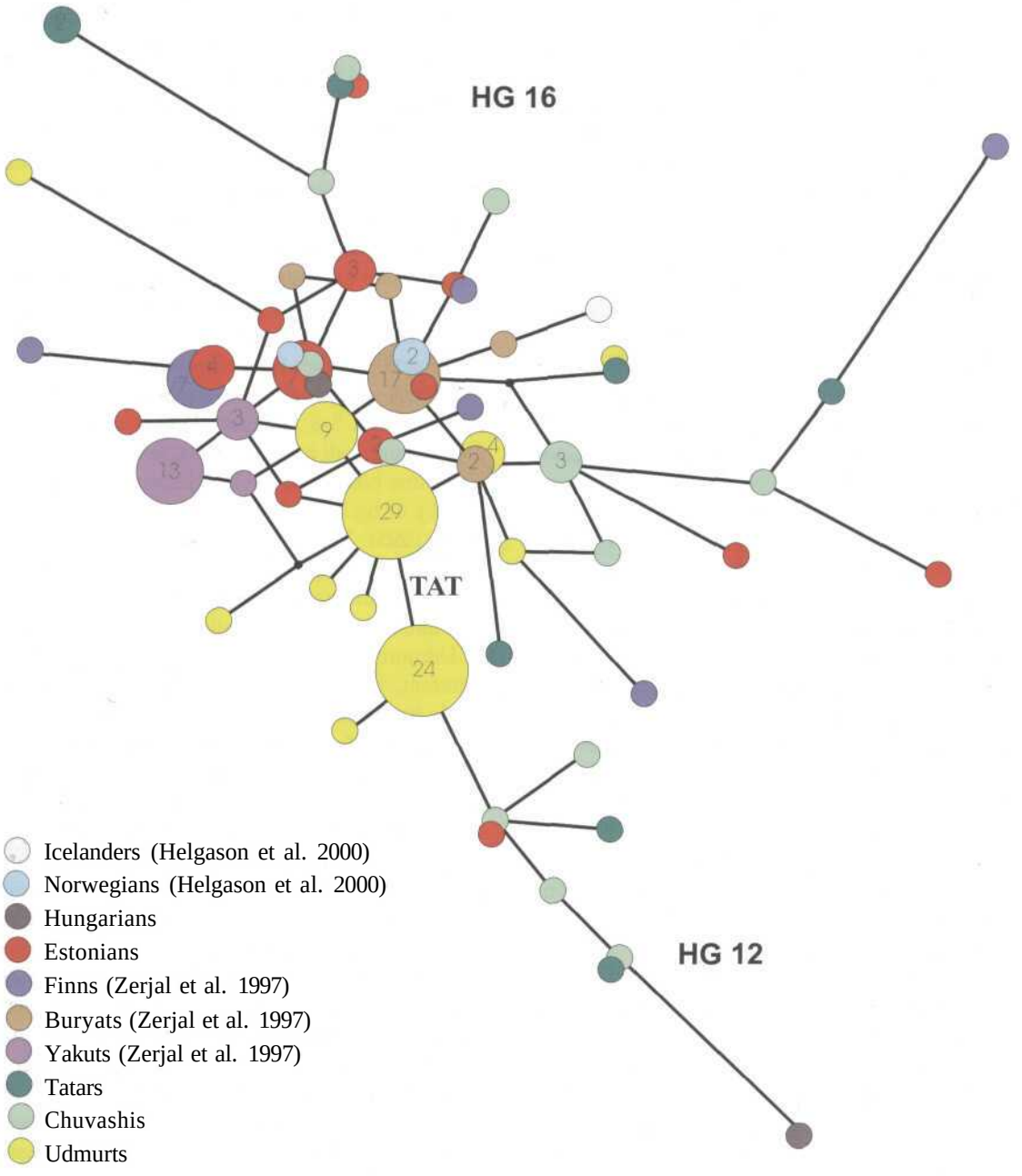

Figure 2. *Median joining network of Y-chromosomal haplogroups 12 and 16 of some Eurasian populations based on haplotypes, constructed from length variants of 5 microsatellites (DYS 19, 390, 391, 392, 393).*

in the paternal lineages of this population. Nevertheless, since the two dominant Udmurti "modal haplotypes" (combined STR haplotypes) are identical in hg 12 and hg 16 (Fig. 2), one can tentatively assume that the ancestral state for hg 16 (Tat C) derives from the dominant Udmurt variety of hg 12. However, because the amount of data for hg 12 diversity is very limited, many other populations should be studied as well until such an important problems can be considered as settled.

The eastern Eurasian share in the eastern European maternal lineage pool.

In present-day Asian populations the mtDNA variation can be classified under two macro-clusters – M and N (Fig. 3). Both of them coalesce to the African macro-cluster L3, which can be considered as the most recent common ancestor of all non-Africans. Hg M and its eastern Asian subhaplogroups (C, D, E, G, Z), each defined by characteristic mutations in mtDNA, comprise about 50 per cent of native Asian mtDNA lineages. The presence of M and its subhaplogroups is very rare in European populations. Contrary to that, macrocluster N is shared by Asians and Europeans. Among them, eastern Asians share mtDNA haplogroups A, Y, B and F: the last two of them derive from the internal node R (Fig. 3).

Most of maternal lineages of Maris, Mordvins, Tatars, Udmurts and Chuvashis belong to western Eurasian mtDNA haplogroups (Table 1). Lineages, characteristic to Asian populations, can be found at frequencies around 10 per cent among Turkic speaking populations, but they are less frequent among Maris and Mordvins. As discussed already earlier (Villems et a1. 1998), the eastern Asiatic component in maternal lineages of the Baltic-Finnic populations is low– less than 1 % among Estonians and Finns (Sajantila et al. 1995; Kittles et al. 1999). On the other hand, geographic distances rather than linguistic background seem to have often played the main role in the development of the pattern of the spread of maternal lineages in Eurasia. The improved data of genetic variation in Volga region populations confirm this observation. It means that the Baltic-Finnic speakers have slightly more similarities with their immediate geographical neighbors in maternal lineages than with linguistic relatives, who live father away. And this slight difference manifests itself as expected – in an increased share of eastern Eurasian lineages in the Volga region. However, one must also remember that the Saami people do have eastern Eurasian variants of mtDNA at significantly higher frequency that either Finns or Estonians – depending on a particular branch of Saami, about 3-8 per cent (Dupuy and Olaisen; 1996, Delghandi et al. 1998; Sajantila et al. 1995). The same holds true for Karelians, where the frequency of hg M is about 5 – 7 per cent (Sajantila et al. 1995). Therefore, one should not over-emphasize the geographic component either, while discussing the genetic relatedness of the European Finnic-speaking populations.

Udmurts, who are also linguistically somewhat different, belonging to Permic branch of Finno-Ugric languages, carry in their mtDNA pool up to 20 per cent of eastern Asian-specific lineages – a highest share among the European Finno-Ugric people studied so far, although still a minority. Meanwhile, more than a half of mongoloid-specific mtDNA lineages among Udmurts belong to hg D, which is spread widely in Mongols (24 %) (Kolman et al. 1996) and native Siberian populations (ca 13 %) (Torroni et al. 1993), but is quite rare in southeastern Asians (Ballinger et al. 1992). Hg Z is also present among Udmurts. This subhaplogroup of M is characteristic to Siberian populations and frequent among Tungusic-speaking populations spread in the northern part of Siberia (Torroni et al. 1993), and present also among Mongols (Kolman et al. 1996) and Turkic-speaking Central Asians (Comas et al. 1998). The same mtDNA control region motif is present also in Saami population (Sajantila et al. 1995). Here, it is worthwhile to remember that more than a half a century ago, an Estonian archaeologist Richard

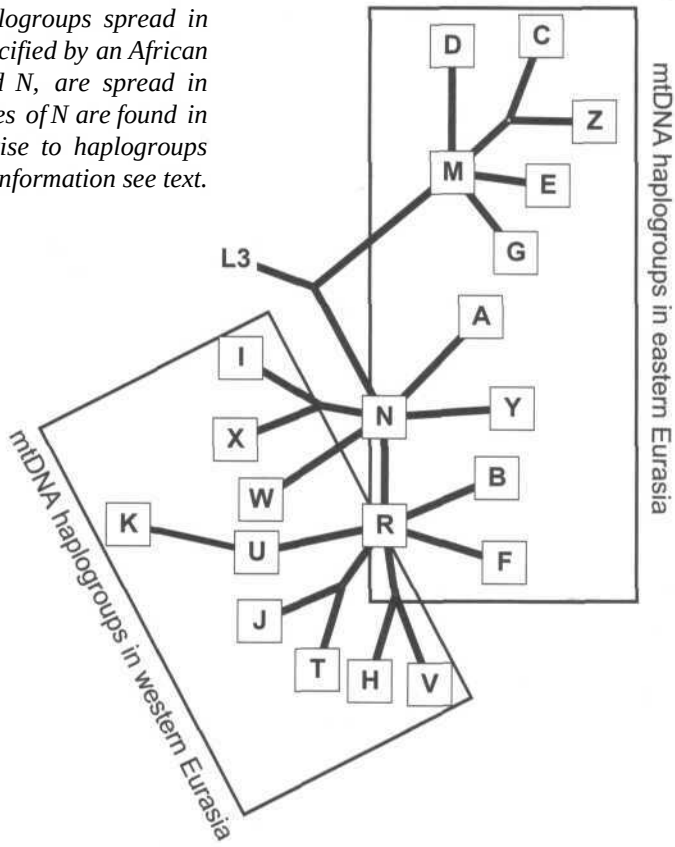

Figure 3. *A tree relating mtDNA haplogroups spread in Eurasia. The direction of the root is specified by an African node L3. Both major branches, M and N, are spread in Asian populations, while only derivatives of N are found in Europe. Further, the N branch gives rise to haplogroups specific to both regions. For additional information see text.*

Table 1. *MtDNA haplogroup frequencies (%) in some European populations.*

	n	H	I	J	K	U	T	V	M+otherEA	other
Finns[1]	123	43.9	4.1	5.7	2.4	22.8	7.3	7.3	0.8	5.7
Estonians	148	50.0	0.7	8.8	2.7	25.7	7.4	1.4	0.0	3.3
Maris	84	40.5	ND	8.3	1.2	22.6	6.0	13.1	4.8	3.5
Mordvins	102	43.1	5.9	6.9	ND	24.5	7.8	3.9	2.9	5.0
Udmurts	90	21.1	ND	2.2	ND	24.4	24.4	ND	18.8	9.1
Tatars	73	35.6	ND	11.0	5.5	15.1	13.7	1.4	10.9	6.8
Chuvashis	55	27.3	1.8	5.5	7.8	36.4	3.6	7.3	9.1	1.2
Bashkirs	40	2.5	2.5	2.5	ND	25.0	2.5	ND	65.0	0.0
Germans[2]	200	52.5	2.5	7.5	7.5	13.5	8.5	3.0	0.0	5.0
Swiss[3]	154	42.9	1.3	11.7	5.2	11.6	16.9	3.2	0.0	7.2
Turks[4]	388	25.0	2.3	10.8	6.2	19.6	11.9	ND	5.2	19.0

Notes:[1] Sajantila et al. 1995, Kittles et al. 1999; [2] Lutz et al. 1998; [3] Dimo-Simonin et al. 2000; [4] Tambets et al. in press;

"n - number of individuals; EA - east Asian haplogroups"

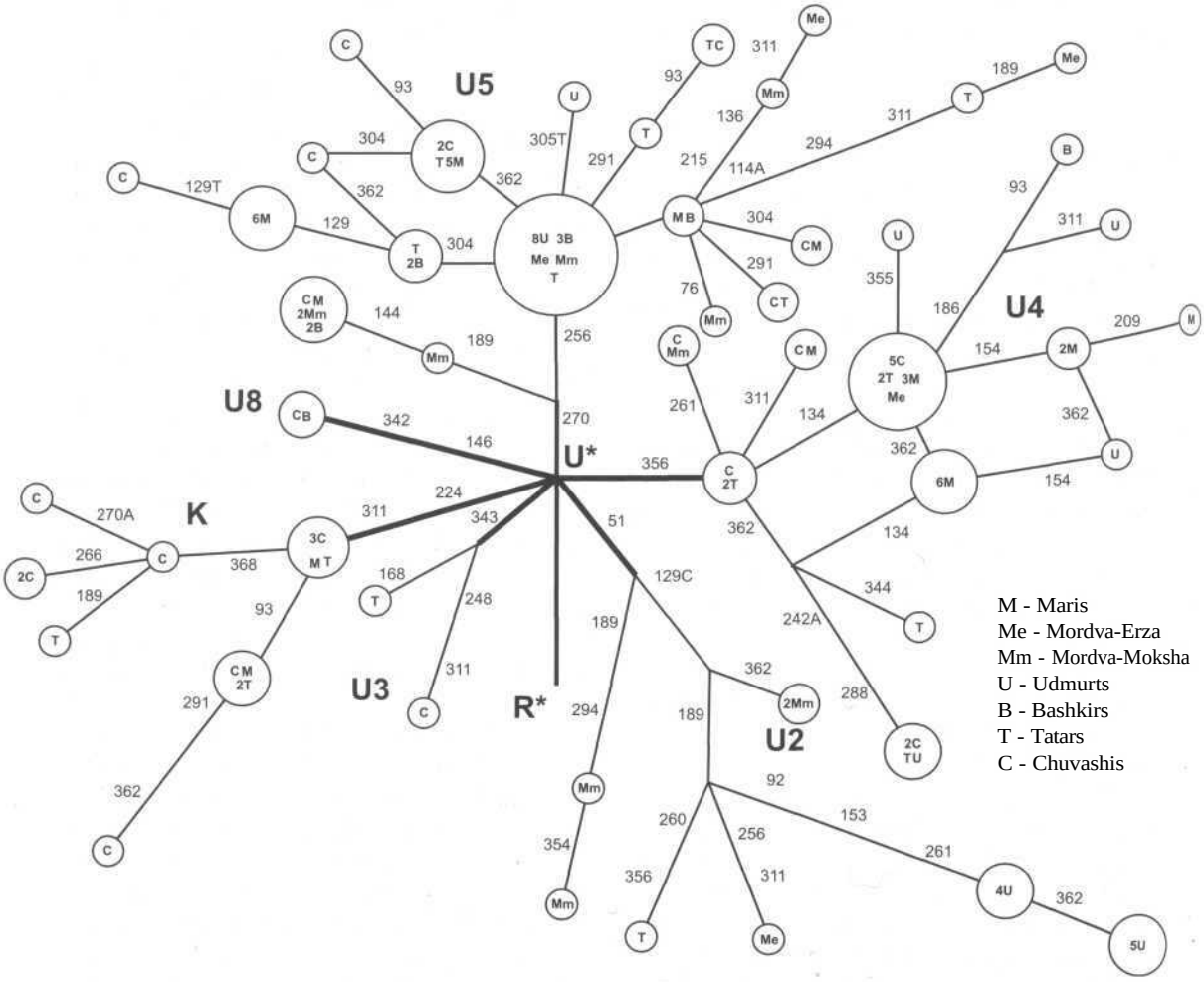

Figure 4. *Median network tree of human mtDNA haplogroup U lineages among some Finno-Ugric and Turkic speaking populations in eastern Europe and Southern Urals. U* - nodal position of hg U, R* - intermediate supergroup, from where most of European lineages originate. Sequence types are represented by circles, with area proportional to frequency. Links are labeled by the nucleotide positions in mtDNA HVS 1 region (minus 16 000) to designate transitions, transversions are further specified.*

Indreko (1948) suggested that at the end of the Upper Paleolithic, a branch of the original Europeans moved from the area of present southern Russia to the Yenisey River and mixed there with original Asian people. He further assumed that one part of this mixed breed moved to the north and back to Europe along the Arctic Zone, eventually contributing significantly to the gene pool of that time northeastern Scandinavians to give rise to the present-day Saami population.

The frequency of haplogroup Z in Udmurts (ca 5 per cent) is very similar to that among Saamis (Sajantila et al. 1995), Mongols (Kolman et al. 1996) and Siberians (Schurr et al. 1999) and slightly lower than among some Central Asian populations (Yao et al. in press). Z is present also among Turkic speaking populations such as Bashkirs and Turks. However, the main specific aspect in Udmurt maternal lineages, irrespective whether eastern or western Eurasian, is in their limited diversity. Because the same finding characterizes their paternal lineages, one may generalize the suggested above conclusion about the role of genetic drift in the demographic history of Udmurts. Again, before going further with conclusions, a more thorough coverage of this population is needed to be sure that some of the phenomena observed are not too much influenced by a sampling effect.

As a further case study, we constructed a phylogenetic tree for the second most frequent in western Eurasia and in India cluster of mtDNA lineages – of hg U (Fig. 4). Quite clearly again, the Udmurts are clustered into a few haplotypes only. The exception is sub-haplogroup U4. This exception is in a way interesting, because this lineage cluster reaches its highest frequency and variability in northeastern Europe and, furthermore, is rare among the Indo-Aryan-speaking people like Iranians, Ossetes and Kurds (an unpublished observation of our laboratory). Yet another specific feature of the Volga-Finnic people can be seen in Fig. 4. Namely, the frequency of haplogroup K (in fact, a variety of hg U) is much higher among the two investigated Volga basin Turkic speaking populations – Chuvashis and Tatars, than among their Finno-Ugric-speaking neighbors. This sharp difference is likely of a "diagnostic value", since hg K is also relatively infrequent in the Baltic Finno-Ugric people and even not sampled in Saami population.

One can, nevertheless, draw a few well supported by new genetic analysis conclusions. The first is that all Volga populations analyzed share, despite of their linguistic differences, an overwhelmingly western Eurasian maternal ancestry. Within this ancestry, one may find differences like an apparent scarcity of haplogroup K among the Finno-Ugric speaking Volga populations, compared to an abundance of this variety of haplogroup U among their Turkic speaking neighbors. However, these are differences within the same general picture. What is a much more basic problem is a question "from where the genetic Asia starts"? And here the answer seems to become now clearer: Bashkirs are the people among whom the eastern Asian maternal lineages are in a substantial majority. The change is very sharp indeed. Whereas their immediate western neighbors and Turkic-speaking linguistic relatives Tatars carry about 36 per cent of the dominant European haplogroup H lineages, then among Bashkirs its frequency drops an order of magnitude (see Table 1). And this example is highly intriguing in a sense that here, both the linguistic and geographic proximity arguments break down, while no apparent geographic barrier can be identified. One may assume that there must be very profound differences in the demographic histories of these two populations. Secondly, irrespective of the observed differences among the Finno-Ugric-speaking Volga populations, they share several, common with the Baltic Finno-Ugric populations' genetic patterns. The most eminent among those is a high frequency of hg 16 of the Y chromosome. But such a deep similarity can be seen also in the scarcity of the maternally inherited hg K. Further research in other FU populations, such as Komis, would tell us whether the observed here phenomena are indeed general.

References

Ballinger, S. W., Schun; T. G., Torroni, A., Gan, Y. Y., Hodge, J. A., Hassan, K., Chen, K. H. & Wallace, D. C. 1992, Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. – Genetics 130, pp. 139-152.

Bandelt, H. J., P. Forster, B. C. Sykes, M. B. Richards 1995, Mitochondrial portraits of human populations using median networks. – Genetics 141, pp. 743–753.

Cann, R. L., M. Stoneking, A. L. Wilson 1987, Mitochondrial DNA and human evolution. – Nature 325, pp. 31–36.

Cavalli-Sforza, L. L., P. Menozzi, A. Piazza 1994, The History and Geography of Human Genes, Princeton, New Jersey.

Cavalli-Sforza, L. L., E. Minch 1997, Paleolithic and Neolithic lineages in the European mitochondrial gene pool. – American Journal of Human Genetics 61, pp. 247-25 1.

Chikhi, L., G. Destro-Bisol, G. Bertorelle, V. Pascali, G. Barbujani 1998a, Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. – American Journal of Human Genetics 95, pp. 9053–9058.

Chicki, L., G. Destro-Bisol, V. L. Pascali, V. Baravelli, M. Dobosz, G. Barbujani 1998b, Clinal variation in the nuclear DNA of Europeans. – Human Biology 70, pp.6434-57.

Comas, D., F. Calafell, E. Mateu, A. Perez-Lezaun, E. Bosch, R. Martinez-Arias, J. Clarimon, F. Facchini, G. Fiori, D. Luiselli, D. Pettener, J. Bertranpetit 1998, Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. – Am J Hum Genet 63, 1824–38.

Delghandi M., E. Utsi, S. Krauss 1998, Saami mitochondrial DNA reveals deep maternal lineage clusters. – Hum Hered 48(2), 108–114.

Dimo-Simonin, N., F. Grange, F. Taroni, C. Brandt-Casadevall, P. Mangin 2000, Forensic evaluation of mtDNA in a population from south west Switzerland. – Int J Legal Med 113, 89–97.

Dolukhanov, P.M. 2000 "Prehistoric revolutions" and languages in Europe. – A. Künnap (ed.), The Roots of Peoples and Languages of Northern Eurasia II and III, Tartu, pp. 71-84.

Dupuy, B. M., B. Olaisen 1996, MtDNA sequences in the Norwegian Saarni and main population. – A. Carracedo, B. Brinkman (eds.), Advances in Forensic Haemogenetics, Berlin, pp. 23–25.

Helgason, A., S. Sigurdardottir, J. Nicholson, B. Sykes, E. W. Hill, D. G. Bradley, V. Bosnes, J. R. Gulcher, R. Ward, K. Stefansson 2000, Estimating Scandinavian and Gaelic Ancestry in the Male Settlers of Iceland. – Am J Hum Genet 67, pp. 697–717.

Indreko, R. 1948, Origin and area of settlement of the Fenno-Ugrian people. – Science in Exile. Publication of the Scientific Quaterly Scholar, Heidelberg, pp. 3–24.

Jobling, M.A., A. Pandya, C. Tyler-Smith 1997, The Y chromosome in forensic analyses and paternity testing. – Int J Legal Medicine 110, pp. 118-124.

Karafet, T. M., S. L. Zegura, 0 . Posukh, L. Ossipova, A. Bergen, J. Long, D. Goldman, W. Kliz, S. Harihara, P. Kniff, V. Wiebe, R. C. Griffiths, A. R. Tempelton, M. F. Hammer 1999, Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. – Am J Hum Genet 64, pp. 817-831,

Kittles, R. A., M. Perola, L. Peltonen, A. W. Bergen, R. A. Aragon, M. Virkkunen, M. Linnoila, D. Goldman, J. Long 1998, Dual origins of Finns revealed by Y chromosome haplotype variation. – Am J Hum Genet 62, pp. 1171–1179.

Kittles, R. A., A. W. Bergen, M. Urbanek, M. Virkkunen, M. Linnoila, D. Goldman, J. C. Long 1999, Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. – Am J Phys Anthropol 108, 381–99.

Kivisild, T., M. J. Bamshad, K. Kaldma, M. Metspalu, E. Metspalu, M. Reidla, S. Laos, J. Parik, W. S. Watkins, 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. – Curr Biol 9, 1331-1334.

Kolman, C., N. Sambuughin, E. Bermingham 1996, Mitochondria1 DNA analysis of Mongolian populations and implications for the origin of New World founders. – Genetics 142, 1321–34.

Krings, M., A. Stone, R. W. Schmitz, H. Krainitzki, M. Stoneking, S. Paabo 1997, Neanderthal DNA sequences and the origin of modem humans. – Cell 90, 19–30.

Krings, M., C. Capelli, F. Tchentscher, H. Geisert, S. Meyer, A. von Haeseler, K. Grossschmidt, G. Possnert, M. Paunovic, S. Pääbo 2000, A view of Neandertal genetic diversity. – Nature Genetics, 26, pp. 144–146.

Lahermo, P., M.L. Savontaus, P. Sistonen, J. Beres, P. de Kniff, P. Aula, A. Sajantila 1999, Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. – Eur J Hum Genet 7, pp. 447–458.

Lutz, S., H. J. Weisser, J. Heizmann, S. Pollak 1998,

Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. – Int J Legal Med 111,67-77.

Malaspina, P., F. Cruciani, P. Santolamazza, A. Torroni, A. Pangarizio, N. Akar, V. Bakalli, R. Brdica, A. Jaruzelska, A. Kozlov, B. Malyarchuk, S. Q. Mehdi, R. Qamar, 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 in press, Patterns of male-specific inter-population divergence in Europe West Asia and North Africa. – Annals Hum Genet.

Niskanen M. 2000 The origin of Europeans: population movements, genetic relationships and linguistic distribution. – A. Kiinnap (ed.), The Roots of Peoples and Languages of Northern Eurasia II and III, Tartu, pp. 33–68.

Orekhov, V., A. Poltoraus, L. A. Zhivotovsky, V. Spitsyn, P. Ivanov, N. Yankovsky 1999, Mitochondrial DNA sequence diversity in Russians. – FEBS Letters 445, 197-201.

Ovchinnikov, I. V., A. Gotherstrom, G. P. Romanova, V. M. Kharitonov, K. Liden, W. Goodwin 2000, Molecular analysis of Neanderthal DNA from the northern Caucasus. – Nature 404, pp. 490–493.

Richards, M., H. Corte-Real, P. Forster, V. Macaulay, H. Wilkinson-Herbots, A. Demaine, S. Papiha, R. Hedges, H. J. Bandelt, B. Sykes 1996, Palaeolithic and Neolithic lineages in the European mitochondrial gene pool. – American Journal of Human Genetics 59, pp. 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, 0. Rychkov, Y. Rychkov, M. Golge, D. Dimitrov, E. Hill, D. Bradley, V. Romano, F. Cali, G. Vona, A. Demaine, S. Papiha, C. Triantaphyllidis, G. Stefanescu, J. Hatina, M. Belledi, A. Di Rienzo, A. Novelletto, A. Oppenheim, S. Norby, N. AI-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, pp. 1251–1276.

Rosser Z H., T. Zerjal, M. E. Hurles, M. Adojaan, D. Alavantic, A. Amorim, W. Amos, M. Armenteros, E. Arroyo, G. Barbujani, G. Beckman, L. Beckman, J. Bertranpetit, E. Bosch, D. G. Bradley, G. Brede, G. Cooper, H. B. S. M. Corte-Real, P. de Knijff, R. Decorte, Y. E. Dubrova, O. Evgrafov, A. Gilissen, S. Glisic, M. Golge, E. W. Hill, A. Jeziorowska, L. Kalaydjieva, M. Kayser, T. Kivisild, S. A. Kravchenko, A. Krumina, V. Kucinskas, J. Lavinha, L. A. Livshits, P. Malaspina, S. Maria, K. McElreavey, T. A. Meitinger, A. – V. Mikelsaar, R. J. Mitchell, K. Nafa, J. Nicholson, S. Norby, A. Pandya, J. Parik, P. C. Patsalis, L. Pereira, B. Peterlin, G. Pielberg, M. J. Prata, C. Previdere', L. Roewer, S. Rootsi, D. C. Rubinsztein, J. Saillard, F. R. Santos, G. Stefanescu, B. C. Sykes, A. Tolun, R. Villems, C. Tyler-Smith, M. A. Jobling 2000, Y-Chromosomal Diversity in Europe Is Clinal and Influenced Primarily by Geography, Rather than by Language. – American Journal of Human Genetics 67, pp. 1526–1543.

Rootsi, S., T. Kivisild, K. Tambets, M. Adojaan, J. Parik, M. Reidla, E. Metspalu, S. Laos, H.-V. Tolk, R. Villems 2000, On the phylogeographic context of sex-specific genetic markers of Finno-Ugric populations. – A. Kiinnap (ed.), The Roots of Peoples and Languages of Northern Eurasia II and III, Tartu, pp. 148–164.

Sajantila, A., P. Lahermo, T. Anttinen, M. Lukka, P. Sistonen, M.L. Savontaus, P. Aula, L. Beckman, L. Tranebjaerg, T. Gedde-Dahl, L. Issel-Tarver, A. DiRienzo, S. Pääbo 1995, Genes and languages in Europe: an analysis of mitochondria1 lineages. – Genome Res 5, 42-52.

Sajantila, A., A.H. Salem, P. Savolainen, K. Bauer, C. Gierig, S. Paabo 1996, Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. – Proc Natl Acad Sci USA 93, 12035–12039.

Schurr T. G., R. I. Sukernik, Y. B. Starikovskaya, D. C. Wallace 1999, Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea – Bering Sea region during the Neolithic. – Am J Phys Anthropol 108: 1–39.

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, 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, pp. 1155-1159.

Simoni, L., F. Calafell, D. Pettener, J. Bertranpetit, G. Barbujani 2000, Geographic patterns of mtDNA diversity in Europe. – American Journal of Human Genetics 66, pp. 262-278.

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. De Stefano, F., R. Villems in press, The Topology of the Maternal Lineages of the Anatolian and Trans-Caucasus Populations and the Peopling of the Europe: Some Preliminary Considerations. – McDonald Institute for Archaeological Research Monograph Series, Cambridge University, Cambridge.

Torroni, A,, R. I. Sukernik, T. G. Schurr, Y. B. Starikovskaya, M. F. Cabell, M.H. Crawford, A. G. Comuzzie, D. C. Wallace 1993, mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. – Am J Hum Genet 53, 591-608.

Torroni, A,, M.T. Lott, M.F. Cabell, Y.S. Chen, L. Lavergne, D.C. Wallace 1994, mtDNA and the origin of Caucasians: identification of ancient Caucasian- specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. – Am J Hum Genet 55,760–76.

Torroni, A., H. – J. Bandelt, L. D'Urbano, P. Lahermo, P. Moral, D. Sellitto, C. Rengo, P. Forster, M. –L. Savontaus, B. Bonne-Tamir, R. Scozzari 1998, mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. – American Journal of Human Genetics 62, pp. 1137–1152.

Torroni, A,, M. Richards, V. Macaulay, P. Forster, R. Villems, S. Norby, M.-L. Savontaus, K. Huoponen, R. Scozzari, H.-J. Bandelt 2000, mtDNA haplogroups and frequency patterns in Europe. – American Journal of Human Genetics 66, pp.1173-1178.

Underhill, P.A., P. Shen, A.A: Lin, L. Jin, G. Passarino, W.H. Yang, E. Kauffman, B. Bonne-Tamir, J. Bertranpetit, P. Francalacci, M. Ibrahim, T. Jenkins, J. R. Kidd, S. Q. Mehdi, M. T. Seilstad, R. S. Wells, A. Piazza, R. W. Davis, M. W. Feldman, L. L. Cavalli-Sforza, P. J. Oefner 2000, Y chromosome variation and the history of human populations – Nature Genetics 26, pp. 358–361.

Villems, R., M. Adojaan, T. Kivisild, E. Metspalu, J. Parik, G. Pielberg, S. Rootsi, K. Tambets, H.-V. Tolk 1998, Reconstruction of maternal lineages of Finno-Ugric speaking people and some remarks on their paternal inheritance. – K. Julku, K. Wiik (eds.), The Roots of Peoples and Languages of Northern Eurasia I, Jyvaskyla, pp. 180–200.

Wallace, D. 1995, Mitochondria1 DNA variation in human evolution, degenerative disease, and aging. – Am J Hum Genet 57, pp. 201-223.

Wiik, K. 2000, European ***lingua francas.*** – A. Kiinnap (ed.), The Roots of Peoples and languages of Northern Eurasia II and III, Tartu, pp. 202–236.

Zerjal, T., B. Dashnyam, A. Pandya, M. Kayser, L. Roewer, F. R. Santos, W. Schiefenhovel, N. Fretwell, M. A. Jobling, S. Harihara, K. Shimizu, D. Semjidmaa, A. Sajantila, P. Salo, M. H. Craford, E.K. Ginter, O. V. Evgrafov, C. Tylor-Smith 1997, Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. – Genome Res 5, pp. 42–52.

Zerjal, T., A. Pandya, F. R. Santos, R. Adikari, E. Tarazona, M. Kayser, O. Evgafov, L. Singh, K. Thangaray, G. Destro-Bisol, M. G. Thomas, R. Qamar, S. Q. Mehdi, Z. H. Rosser, M. E. Hurles, M. A. Jobling, C. Tyler-Smith 1999, The use of **Y**-chromosomal DNA variation to investigate population history: Recent male spread in Asia and Europe. – S. Papiha, R. Deka, R. Chakraborty (eds.), Genomic Diversity: Application in Huinan Population Genetics, New York, pp. 91–101.

Yao Y.-G., X.-M. Liu, H.-R. Luo, W.-H. Li, Y.-P. Zhang in press, Gene admixture in the Silk Road region of China – evidence from mtDNA and melanocortin 1 receptor polymorphism. – Genes Genet Syst.

HISTORICA FENNO-UGRICA

The Roots of Peoples and Languages of Northern Eurasia IV

Oulu 18.8.–20.8. 2000

Edited by Kyosti Julku

Editorial Assistants
Merja Äarela
Eila Kortet

Societas Historiae Fenno-Ugricae
Oulu 2002