TRAMES, 2001, **5**(**55**/**50**), 1, 59–74

THE CONCEPTS OF RICHARD INDREKO ABOUT THE ORIGIN OF THE FINNO-UGRIC SPEAKERS AND THE POPULATION GENETICS OF THE EXTANT NORTH-EAST EUROPEAN POPULATIONS

Kristiina Tambets, Siiri Rootsi, Toomas Kivisild, Richard Villems

University of Tartu and Estonian Biocentre

Abstract. Archaeologists define their findings in terms of cultures and industries, linguists operate with languages, and the basic unit in biological anthropology is population. Intermarriages of these three systems have raised, and are still doing so, fascinatingly controversial speculations. In the current paper our intention is to discuss the accumulating data of population genetics in a specific context: in the light of scenarios of Richard Indreko, suggested in his half a century old paper. Our approach is to couple the genetic evidence (mitochondrial DNA and Y chromosome data) of Finno-Ugric speaking populations with this archaeological vision.

Introduction

Anthropological data concerning the populations speaking Finno-Ugric languages started to emerge in the 19th century. Although the first studies were almost anecdotal in their methods and conclusions, they regretfully continued to be influential in one aspect: in speculating about the mongoloidness in the genetic substratum of the Finno-Ugrians. Namely, an acknowledged French anthropologist Paul Broca deduced from three crania that Estonians, together with the Finns, belong to the Mongoloid race. Although later much more detailed anthropometric studies (Aul 1964, Mark 1956) have disputed such an interpretation, the notion of Asian origin of the Finno-Ugric people has found its way from textbook to textbook and the idea surfaces now and then in recent articles as well (Zerjal et al. 1997; Zerjal et al. 1999).

From the eighties onwards, the breakthrough in DNA sequence analysis paved the way for molecular anthropology, the potential of which lies now in the abundance of possible markers that can be used in phylogenetic studies. Selectively neutral markers that lie within the mitochondrial DNA (mtDNA) or Y

chromosome have become particularly popular because they are inherited uniparentally, i.e. only maternally or paternally. They thereby allow establishing lineages of individuals in a similar fashion that family trees are constructed. The concert of such lineages collected from a population can therefore have an internal or external most recent common ancestor (MRCA) in respect to other populations. Furthermore, applying the steadiness of molecular clock (the number of randomly accumulated mutations is nearly constant in time), coalescence time can be calculated for all the lineages or those of particular interest.

From the broader circle of questions being addressed in the paper by Richard Indreko (1948) we have chosen the following few basic concepts, and analyse their concordance with the evidence available from mtDNA and Y-chromosomal studies. Here are the statements taken from Indreko (1948) and formulated by us as questions:

1. The earliest, pre-Neolithic layer of human population in eastern Baltic and Finland was established by the descendants of Palaeolithic Europeans from western and central Europe who followed the receding ice border together with the reindeer.

Question 1. Can we find lineages among the extant Finno-Ugric populations that can be traced back to common European founder motifs within a period of 10,000 to 60,000 years? The latter estimate is the maximum time limit, according to current archaeological evidence, for the entry of modern humans to Europe.

2. At the end of Upper Palaeolithic, a branch of the original Europeans moved over the present southern Russia to the upper course of Yenisei River and mixed there with original Asian people. This mixed population spread further northwards to the Arctic Zone and thence back west up to Scandinavia. There this branch intermixed again with local northern populations, the result of which can still be recognised in extant Laplanders (Saamis) and Samoyeds.

Question 2. Do we find traces of lineages of Asian origin among the present day Saamis that do not exist among other Finno-Ugrians, or among Indo-Europeans in Western Europe either?

3. The Neolithic Combed Ware culture developed locally from that of the original Europeans and belongs, though not fully, to the Finno-Ugrians. The Asiatic centre of culture showed considerable deviations, and no communications existed with the woodland zone of Eastern Europe. Because of the improvement of environmental conditions, there was a general tendency of the proto-Finno-Ugric populations to spread towards northern and eastern territories (Fig. 10, 11 in Indreko 1948 see the present issue, p. 18).

Question 3. According to this point, we should not find any lineage clusters specific to the Finno-Ugrians older than roughly 10,000 years. The southern- and westernmost populations should then also have greater variability than the northern and eastern populations, derived from them. Are the genetic lineages of Finno-Ugrians then just a sub-section of the genetic pool intrinsic for all Europeans in general with only a few population-specific polymorphisms appearing on tips of the western Eurasian genealogies?

4. Towards the end of the Stone Age the carriers of the Boat Axe culture pressed forward from central Germany towards north-eastern Europe bearing with them the Indo-European languages. In the eastern Baltic and Finland these people intermixed with the existing Finno-Ugrians. Finally, in the northern Baltic area the Finnic languages prevailed, while on the southern and western shores of the Baltic Sea the Indo-European acquired authority.

Question 4. The time window between the postulated migrations taking place in the Baltic region during the early and late Neolithic is too narrow for accumulating detectable genetic differentiation in the area. Assuming, however, that the two migrations originated from populations that were already different, at least in the frequencies of some lineages, the gradual dilution of the new wave of migration against the background of the already existing populations could then have left an imprint, now recognisable in a steady decrease of some specific lineages for the Balto-Finnic populations in south-west and, respectively, an increase of variants that are more abundantly spread around central Germany. Can we define such differences and establish whether the meeting of these Neolithic populations has left behind any genetic sign?

To find answers to these questions associated with the population history of the Finno-Ugric speaking populations, one will need to compare them against the general background of other Eurasian populations. First we analyse our female lineages through the mtDNA variation and then the history of paternal bloodline by comparing markers from Y chromosome. In both cases we cannot bypass the method:

Molecular family trees

In the case of classical genetic markers and in anthropometrical or anthroposcopical data sets, the units under study are usually statistical averages of the populations. In contrast, the uniparentally inherited mtDNA and Y-chromosomal markers enable to operate with real genealogies of individuals taken from those populations, and the final outcome is a truly cladistic interpretation. The classical tree building methods have proved to be futile in mainstream population genetics, where rapidly evolving mtDNA and Y-chromosomal microsatellites are being used. Fast evolving sites introduce incompatibilities resulting in multiple equally parsimonious trees. One possible solution to this problem has been proposed: to use networks instead of trees (Bandelt et al. 1995). A network can contain millions of equally parsimonious trees, displaying them in the form of reticulations, where the nodes of the network correspond to specific *haplotypes*, defined by specific diagnostic sites or mutations in genome, which are given as links between two nodes. An additional benefit is that the genealogical approach permits one to group multiple different *haplotypes* into a limited number of *haplogroups* and thus to systemise the data in a cladistic way. The nomenclature of *haplogroups* used here is according to Torroni et al. (1996) for mtDNA and according to Jobling et al. (1997) for Y chromosome.

Maternal evidence

Planetary context

We harbour mitochondria in our cells alongside other eucaryotes already a billion years or more. All the mitochondria co-evolving with the present-day human species can be traced back to the MRCA, i.e. a mitochondrion from a single woman who probably lived around 100,000-200,000 years ago in Africa (Cann et al. 1987) among at least 2,500 other healthy women of our species. After the split of African and non-African populations about 50,000 to 100,000 years ago our grannies moved apart into separate geographical environments. Partial isolation of continental population groups has now resulted in highly discriminative pattern of mtDNA variation on global scale. Figure 1 sums up recently accumulated evidence in this perspective. For the purposes of our current quest, the variation in Eurasia is shown in detail, while the highly diverse African clusters are not specified. Haplogroups H, U, T and J characterise the major portion (more than 80%) of European mtDNAs, whereas haplogroups A, B, F and M dominate in Asia. There is a common overlap of less than 5% between the populations from different continents except for zones of intermixture, e.g. Central Asia (Comas et al. 1998). The general division of mtDNA lineages into eastern and western Eurasian sets is currently best explained by at least two different dispersals of modern humans, originating from eastern Africa (Lahr and Foley 1994). The earlier one used the southern route and brought forward the proto-Asian lineages (Quintana-Murci et al. 1999), while the later migration, possibly by the northern route, gave rise to the western Eurasian-specific haplogroups (Kivisild et al. 1999).

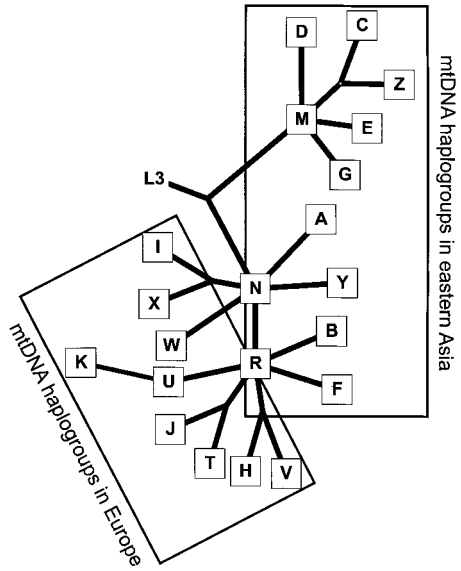

Fig. 1. A tree relating mtDNA haplogroups spread in Eurasia. The root is specified by an African node L3. Both major branches, M and N, have spread in Asian populations, while only derivatives of N are found in Europe. Further, the N branch gives rise to haplogroups specific to both regions.

Maternal lineages of Finno-Ugric speaking populations in European context

Limiting ourselves now to the analysis of mtDNA variation among the Finno-Ugrians, we require data to contrast them with populations from a broader range that would still be comparable in detail. Figure 2 (on the insert between p.64, 65) comprises an analysis of a composite of Finno-Ugric populations together with a

similar but linguistically heterogeneous composite from western Europe. The same set of basic western Eurasian specific lineage clusters characterises the populations under observation. The basic topology of the European mtDNA network would not change significantly if any of these given populations are replaced with some other sufficiently large European population studied so far. It means that this common set of maternal lineages divides relatively uniformly between different populations, regardless of their language affiliation. To provide this statement with quantitative evidence, Table 1 compares frequencies of mtDNA haplogroups of Finno-Ugric populations with some other European populations. The most important aspect at this stage is the uniformity of the distribution of the frequencies of the eight basic western Eurasian-specific haplogroups. Although there are some variations in frequencies, all the major haplogroups are found in eastern as well as in western European populations; among the Basque, Indo-European speakers and Finno-Ugrians.

Table 1.

mtDNA haplogroup frequencies (%) in European populations

	n	H	I	J	T	U	K*	V	W	X	L	M	other
Estonians[1,2]	148	50.0	0.7	8.8	7.4	23.0	2.7	1.4	3.4	0.7	0.0	0.0	2.0
Finns[2,3]	123	43.9	4.1	5.7	7.3	20.4	2.4	7.3	6.5	0.8	0.0	0.8	0.8
Karelians[2]	82	42.7	1.2	3.7	7.3	28.1	2.4	6.1	3.7	0.0	0.0	0.0	4.9
Hungarians[4]	116	41.0	1.7	7.8	11.0	17.3	4.3	2.6	4.3	0.9	0.0	2.6	6.8
Slavs[5,6]	224	39.3	2.1	11.3	13.3	17.9	4.1	2.2	1.2	1.1	0.0	1.5	6.3
Norwegians[7]	216	51.9	2.3	6.5	7.9	16.2	5.6	4.2	1.4	0.5	0.0	0.9	2.8
English[8]	100	51.0	2.0	11.0	6.0	10.0	10.0	3.0	0.0	2.0	1.0	1.0	3.0
Germans[9]	200	52.5	2.5	7.5	8.5	13.5	7.5	3.0	1.0	0.5	0.0	0.0	3.5
French[10,11]	135	45.9	0.7	6.7	11.9	15.6	7.4	3.0	1.5	1.5	2.2	1.5	2.2
Italians[12,13]	147	36.1	4.1	9.5	9.5	18.4	7.5	3.4	2.0	4.8	0.7	0.7	3.4
Spanish[13]	182	45.1	1.1	4.4	9.9	14.3	3.3	5.5	1.1	1.6	3.3	0.5	9.9
Turks[14]	388	25.0	2.3	10.9	11.9	19.1	5.9	0.3	3.9	4.4	0.3	4.1	11.9

[1]Villems et al.1998, [4]Tambets et al. manuscript in preparation, [5]Kivisild et al. 1999, [11]Laos et al. manuscript in preparation, [12] Torroni et al. 1996, [13]Torroni et al. 1997, [14]Tambets et al. in press;
deduced from sequence data: [2]Sajantila et al. 1995, [3]Kittles et al. 1999, [6]Orekhov et al. 1999, [7]Opdal et al. 1998, [8]Piercy et al. 1993, [9]Lutz et al. 1998, [10]Rousselet and Mangin 1998,

* K is a subhaplogroup of U.

The majority, *ca* 80% or more, of the existing European maternal lineages belong to haplogroups H, I, T, U, V, X and are regarded as descendants of the European Upper Palaeolithic gene pool as far as the accumulated diversity within these groups coalesces to 13,000-50,000 years BP. The oldest among them is haplogroup U, or particularly one branch of it – U5, the coalescence age of which dates back to the times when Europe was first colonised by modern humans (Richards et al. 1998). Only a minor part, represented by certain sub-founders from haplogroups J, K and T1, is thought to have been carried to Europe later on by Neolithic farmers (Metspalu et al. 1999, Richards et al. 1996, 1997, 1998).

Coming now back to **the first question** about the place of Finno-Ugric genes in the general Eurasian context, mtDNA data suggest that maternal lineages of the Finno-Ugric speaking populations are composed of the same 8 basic western Eurasian-specific haplogroups that reflect the common mtDNA pool originating from the late Pleistocene and Holocene migrations to Europe. In that general level, the extant Finno-Ugric populations are not distinguishable from other European populations.

What we must bear in mind is that the uniformity of European mtDNA pool is the one we have reached through the first approximation. A bit higher resolution – "unravelling" haplogroups by phylogenetically defined subdivisions – allows to go further and seek for the otherwise hidden but sometimes very informative connections between maternal lineages of extant populations. Let us take haplogroup U as an example. Figure 3 shows quite clearly that the distribution of the sub-clusters of U is far from uniform. Sub-clusters U4 and U5 are notably more frequent among north-eastern Europeans, including Finno-Ugrians, whereas U1, U3 and K have higher frequencies in south. Hence, a more detailed analysis starts to reveal patterns of variation, which splits the descendants of the common founders of the Caucasoid maternal lineages and reveals their subsequent radiation. There is a great demand for additional data for further analysis. Interestingly, one particular population among Finno-Ugrians has always been considered as an exception.

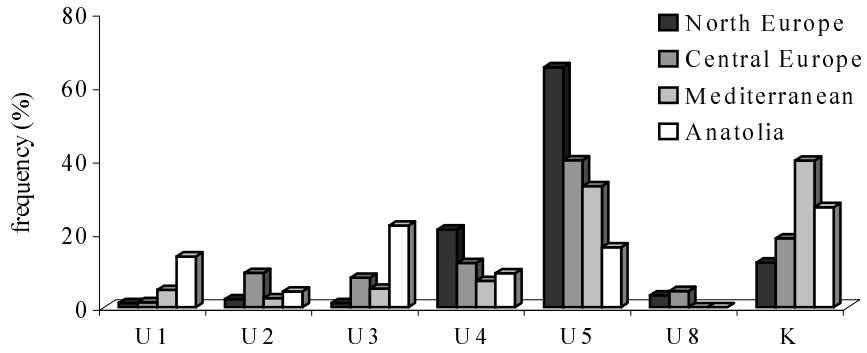

Fig. 3. Geographical distribution of sub-clusters of haplogroup U.
North Europe – Estonians[1,2], Karelians[2], Finns[2,3]; Central Europe – Hungarians[4], Slavs[5,6]; Mediterranean – French[10,11], Spanish[13], Italians[12,13], Sicilians[15]; Anatolia – Turks[14], Armenians[14], Georgians[14], Greeks[15].
The references are numbered as in Table 1, and [15] – Metspalu et al. in preparation.

Saamis – genetic "outliers" or kins?

Genetic distance estimates based on classical genetic markers (Cavalli-Sforza et al. 1994), as well as mtDNA data (Sajantila et al. 1995; Lahermo et al. 1996) have shown very convincingly that Saamis are quite distinct from other European

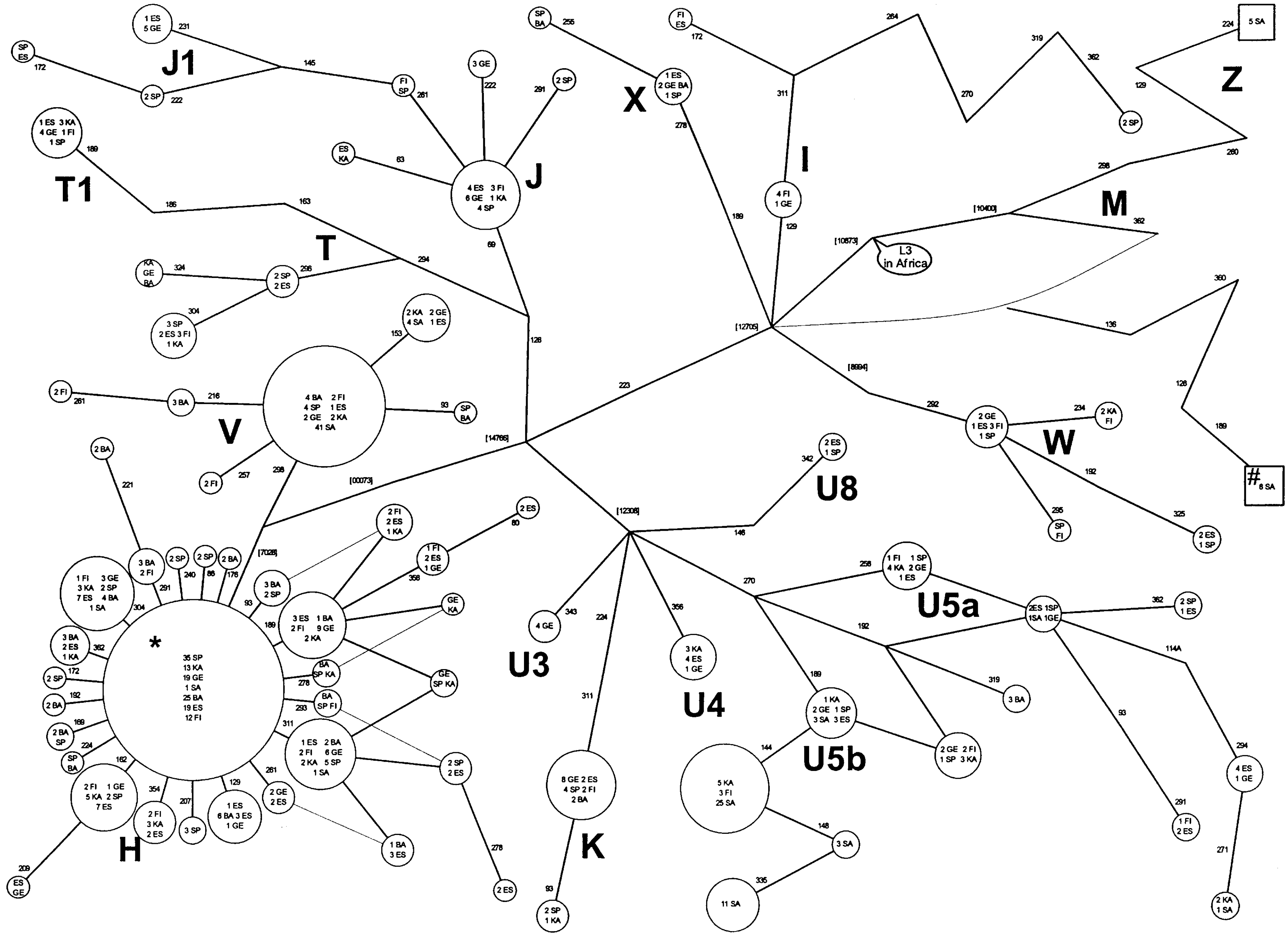

Fig. 2. Phylogenetic reconstruction of mtDNA lineages of Finno-Ugric populations against the background of western European populations. SA – Saamis (Sajantila et al. 1995); ES – Estonians (Villems et al. 1998; Sajantila et al. 1995); FI – Finns (Sajantila et al. 1995; Richards et al. 1996); KA – Karelians (Sajantila et al. 1995); BA – Basques (Bertranpetit et al. 1995; Richards et al. 1996); GE – Germans (Richards et al. 1996); SP – Spanish (Corte-Real et al. 1996; Richards et al. 1996; Salas et al. 1998). The reconstruction is based on hypervariable region 1 sequences and haplogroup assignments are based on characteristic coding region sites (in parentheses) tested on Estonian sample. Only haplotypes represented by two or more individuals are shown. Haplotypes belonging to major western Eurasian-specific haplogroups (Torroni et al. 1996; Macaulay et al. 1999) are shown in circles while those belonging to Asian-specific haplogroup M in rectangles. The place of the reference sequence (Anderson et al. 1981) in the phylogeny is indicated by an asterisk. Dashed line from where a Saami haplotype (#) originates indicates the uncertainty of its haplogroup affiliation. The connection point (root) of the Eurasian lineages clusters to the African mtDNA phylogeny through the consensus of haplogroup L3 is shown in a bubble.

populations. Their genetic distance from other Europeans is much larger than typically between different Europeans, including Finns (Sajantila et al. 1995; Lahermo et al. 1996). These studies have fuelled the idea that Saamis do not have much in common with their close neighbours, leading even to the hypothesis of a language replacement in Finns (Sajantila and Paabo 1995).

A different conclusion can be drawn when searching for phylogenetic explanations for the genetic distances of the Saamis (Villems et al. 1998). The lion's share of the Saami mtDNA lineages, as shown in Fig. 2, splits into two typical European-specific branches – U5b and V – which are spread among Estonian, Finnish, Mari, Hungarian, as well as among other European populations, yet not in eastern Asians. The reason of large genetic distances from neighbours lies in the severely restricted mtDNA pool of Saami population, which has possibly developed due to a very strong random genetic drift, including bottleneck phenomena and founder effects. Only a small part (< 10%) of the Saami maternal lineages is derived from eastern Eurasian gene pool. Although the eastern Asian-specific lineages have been sampled occasionally in other European populations (at frequencies below 1–2%), the haplogroup Z lineages of the Saami are particularly interesting. This haplogroup, recently defined by Schurr et al. (1999), is one of the subgroups of haplogroup M among Siberian populations and is most frequent among Tungusic-speaking populations spread in northern Siberia. Thus, the presence of Z among Saamis could give **an answer to the second question** we set above, being, perhaps, a genetic trace of the mixture of those Europeans who reached the Yenisei River, got blended there and then turned back west over the Arctic Zone, as suggested by Indreko (see above).

From the East or to the East?

Early in this century, Setälä (1926) formulated a theory, according to which the "original home" of Finno-Ugric language speakers was situated somewhere near the shores of the Middle Volga and the Ural Mountains, where they all lived together up to 2 500 BC. First the Ugrian branch, then around 1 000 BC also the Permian branch and during the last centuries of the first millennium BC the Balto-Finnic set off on their way to west. Based on archaeological findings (the spread of Combed Ware culture), the theory about any restricted area as a place of common origin has received much criticism and has been regarded as too superficial. Instead, it has been suggested that proto-Finno-Ugrians may have covered, although sparsely and irregularly, the entire area between the Baltic Sea and the Ural Mountains (Indreko, 1948).

As we already saw, the rarity or lack of the eastern Asian varieties in their mtDNA lineages contradicts the theory of Siberian origin of Finno-Ugric ancestors, and we cannot see any traces of a massive migration of the Finno-Ugric speakers´ maternal lineages from East: these are much the same among both the Estonians and Finns from the shores of the Baltic Sea and among Maris and Mokshas from Volga region (Sajantila et al. 1995). On the other hand, mtDNA data do not contradict the idea that the large area between the Baltic and the Urals

could have been populated by proto-Finno-Ugric people. The coalescence ages of these mtDNA haplogroups lineage clusters that are present among the extant Finno-Ugric populations, are often much older than 10,000 years and coalesce together with those present among other Europeans. So the **answer to the third question is** basically the same as to the first one: the source population for extant Finno-Ugrians and other Europeans is the same and the migration of our ancestors has taken place rather *to the East* than *from the East*, according to the maternal lineages.

The existing mtDNA data do not provide an answer to **the fourth question**, partly because the analyses of maternal lineages of some parts of Baltic region are still incomplete – the data from Latvia and Lithuania are particularly scarce – and more importantly because of the time limits between different migrations: they were too short to accumulate distinguishable genetic differences, in case these waves of migration can be ascribed to relatively recently divided populations.

The spread of maternally and paternally inherited genes in humans may differ significantly. To find out possible connections between archaeological and genetic data we should consider also the paternal aspect of the topic.

Paternal family tree

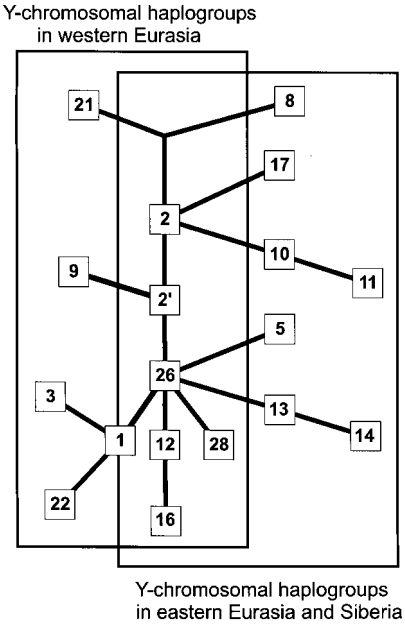

Fig. 4. An unrooted tree relating Y-chromosomal haplogroups, adapted from (Tyler-Smith 1999), spread in Eurasian populations at frequencies more than 2%.

Figure 4 shows the spread of Y-chromosomal haplogroups in Eurasian populations. Most of the paternal lineages found in Estonians as well as among Finns (Lahermo et al. 1999) and Hungarians are those present in other European and western Eurasian populations (see Table 2). Among them, more frequently detected haplogroups are 3 and 1. At lower frequencies also haplogroups 2', 12 and YAP chromosomes of haplogroup 21 are found. The share of Y-chromosomal haplogroups between Finno-Ugric speakers and other European populations is in concordance with the results discussed above in the mtDNA perspective, answering the questions about the earliest European background in the gene pool of Finno-Ugrians.

As we can see, the central nodes: 2, 2' and 26 overlap in Europeans and Asians and therefore the phylogenetic resolution is still insufficient. Clearly, new informative markers are needed. Nevertheless, there are

markers that cast light on the ethnogenesis of Finno-Ugric speaking populations, and we are going to discuss them below.

Table 2

Y chromosomal haplogroup frequencies (%) in European populations

	n	1	2	2'	3	9	12	16	21	26	28
Estonians[1]	167	9.6	0.0	9.6	27.5	1.2	4.8	37.7	3.6	6.0	0.0
Hungarians[1]	113	24.8	2.6	31.0	18.6	15.0	0.9	0.9	6.2	0.0	0.0
Russians[1]	96	7.3	0.0	17.7	47.9	1.0	5.2	13.6	6.3	1.0	0.0
Western-Slavs[2]	216	18.5	0.0	18.5	46.7	5.1	2.3	1.8	6.0	0.9	0.0
Trans-Caucasians[2]	200	27.0	0.5	31.5	5.0	28.5	0.0	1.5	3.0	1.5	1.0

[1] Rootsi et al. 2000
[2] Adojaan et al. manuscript in preparation

The Tat C allele and the origin of the Nordic people

One particularly interesting marker in Y chromosome in the Finno-Ugric context is a T to C point mutation in RBF5 locus, called *Tat* C allele. According to the nomenclature of Jobling et al. (1997) it is haplogroup 16. The first paper on this subject was published by Zerjal et al. (1997), where it was shown that the spread of the *Tat* C allele is restricted mainly to the speakers of two language families: Altaic (Buryats and Yakuts) and the Finno-Ugric branch of the Uralic (Saamis and Finns). Among these populations, this haplogroup was found to be very frequent, covering about half or more of their Y chromosomes. The authors suggested that the Y-chromosomal heritage of Finno-Ugric speaking populations strongly supports an old theory of the Siberian common ancestry of these populations. The main reason why they thought so was because they found the frequency of *Tat* C among Yakuts and Buryats to be higher than among Finns and Saamis, whereas the diversity of microsatellites in the background of *Tat* C carriers among the Baltic Finns and the Siberian populations was nearly equal.

The direction of the spread of Tat C

As it turned out, it is not entirely true. We and others have now studied the microsatellite length variation inside haplogroup 16 in many populations. The values of genetic diversity were calculated by the method of Nei (1987) and are shown in Figure 5. It turned out that among European Finno-Ugric populations, particularly among Estonians, the microsatellite diversity in *Tat* C carriers is higher than in the examined Siberian populations, where its value is extremely low. Equally high diversity of *Tat* C allele among Russians as well as Volga-basin populations Chuvashis and Tatars supports our suggestion that the Eastern European Plain is the “homeland” of this particular variety of the Y chromosome (Rootsi et al. 2000). Therefore, it is reasonable to suggest that the likely paternal gene flow of *Tat* C variant was not from east to west, but from west to east.

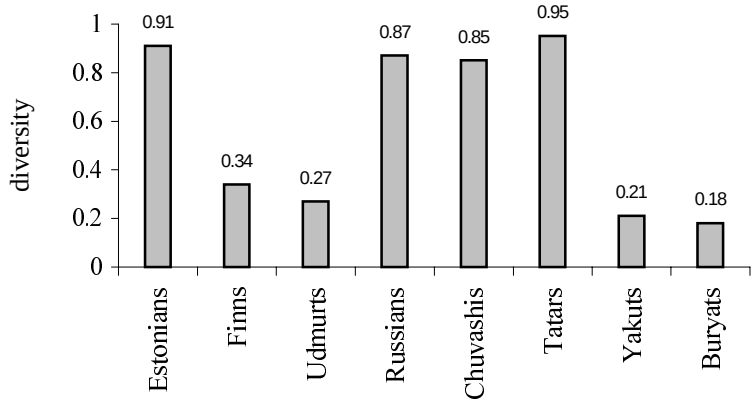

Fig. 5. *Tat* C-allele diversity among some European and Siberian populations.
Indicated populations: Finns, Yakuts, Buryats – Zerjal et al. 1997; others – our data and Rootsi et al., in press.

Geographical distribution of Tat C: further details

Zerjal et al. (1997) have shown that *Tat* C is frequent in Buryats, but nearly absent in Mongols and is not found further south in China, Korea, etc. They also demonstrated a decreasing frequency of *Tat* C among Indo-European language speaking populations: down to 5% in Norwegians and plummeting to zero in Western Europe. Several labs (we, Lahermo et al. 1999, Karafet et al. 1999) have carried out denser mapping of *Tat* C in many other regions of the world (see Fig. 6). These results confirmed the earlier data and brought out new interesting details, as well as posed new questions. *Tat* C was found to be also frequent among Komis, Finno-Ugrians of the Volga-basin and in western Siberia. Furthermore, it was found to be frequent not only in Yakuts and Buryats but also among Mansis, Khanties, Koryaks, Chukchis, Evenks, Nenets, Yukaghirs and Greenland Inuits. In respect of the presence and the frequency (30%) of *Tat* C, which is the only marker studied so far in Saami populations, they resemble other Finnic populations.

In addition to Estonians we have by now investigated several different Slavic and other European populations: Russians, Slovaks, Czechs, Poles, Croats, Lithuanians, Hungarians, Turks, Udmurts, Tatars, Chuvashis and Trans-Caucasus region populations such as Georgians, Armenians, Ossetes. These results allow to draw several important conclusions. First, none of the western Slavic populations (Slovaks, Czechs, Poles, Croats) sampled so far possess *Tat* C at frequencies above a few per cent at best. This confirms our earlier suggestion that a relatively high incidence of *Tat* C among Russians (about 15%) probably reflects a Finno-Ugric "substratum" in eastern Slavs (Villems et al. 1998). The absence or very low incidence of *Tat* C in the Caucasus area shows that paternal lineages of the population of the Eastern European Plain had hardly contributed into the Y-chromosomal pool of the Trans-Caucasians.

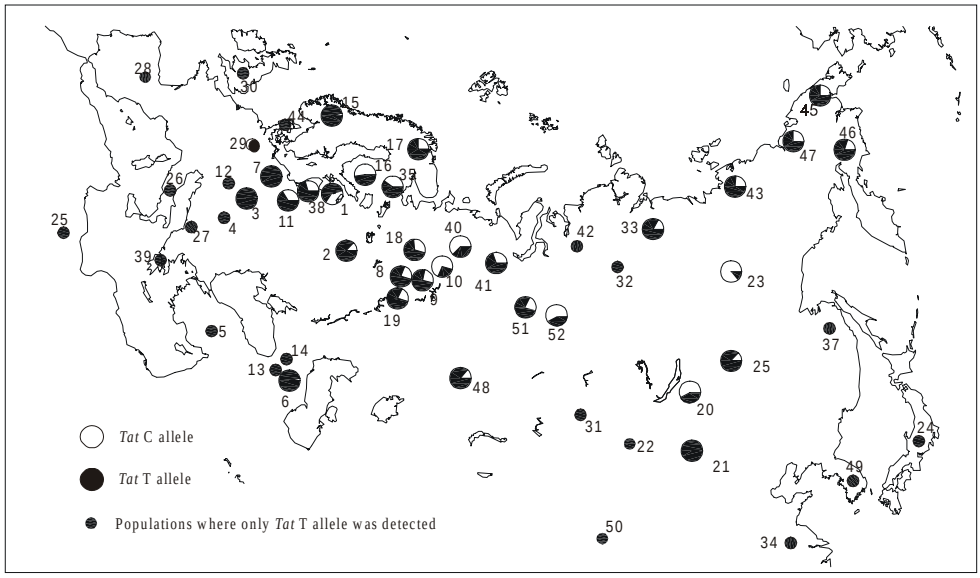

Fig. 6. Geographic distribution of 52 Eurasian populations investigated for *Tat* polymorphism.
*Indicated populations:*1. Estonians 2. Russians 3. Slovaks 4. Hungarians 5. Turks 6. Armenians 7. Poles 8. Chuvashis 9. Tatars 10. Udmurts 11. Lithuanians 12. Czechs 13. Georgians 14. Ossetes 15. Norwegians 16. Finns 17. Saamis 18. Maris 19. Mordvins 20. Buryats 21. Khalks 22. Mongolians 23. Yakuts 24. Japanese 25. Algerians 26. Italians 27. Albanians 28. Basques 29. Germans 30. British 31. Altais 32. Ketis 33. Evenkis 34. Chinese 35. Karelians 36. Koryaks 37. Nivkhis 38. Latvians 39. Greeks 40. Komis 41. Nenets 42. Selkups 43. Evens 44. Danish 45. Chukchis 46. Koryaks 47. Yukaghirs 48. Kazakhs 49. Koreans 50. Tibetians 51. Mansis 52. Khanties;
(Populations 1...14 – our data and Rootsi et al., in press; 15...34 – Zerjal et al. 1997; 35...38; 51…52 – Lahermo et al. 1999; 39...50 – Karafet et al. 1999)

The situation with Latvians and Lithuanians is of particular interest, because they are Indo-European speakers and linguistically close to the Slavs. It turned out that the frequencies of *Tat* C allele in both Latvians (Lahermo et al. 1999) and Lithuanians (our results) are close to that among Estonians, Karelians and Finns: about 29% for Latvians and 33% for Lithuanians, respectively. It is significantly higher than among Russians and other Slavic populations. This finding can be considered interesting from the point of view of the ethnogenesis of the extant Baltic and Finno-Ugric populations: the north-south gradient of *Tat* C allele from the Arctic Sea to Lithuania is insignificant but there is a sharp east-west cline both in Scandinavia and in the Baltic area (Fig. 6). We stress that such a discontinuity is not seen in the distribution of maternal lineages in Europe.

Returning to the third and fourth question asked previously: a possible explanation to the sharp cline in the *Tat* C frequency between Lithuanians and Poles may be connected to the carriers of comb-ceramic culture (Indreko, 1948). By the opinion of some arhaeologists, the southern boundary of the comb-ceramic culture stretched from the mouth of Vistula River across northern Poland, across

the upper course of the Dniepr, across the Don to the Oka River and then turned from the middle of the Volga to the Ural Mountains in the period of about 7 000…3 500 BC (Indreko 1948). The western border of the spread of the Combed Ware culture overlaps with the western and southern borders of the *Tat* C allele distribution in Europe. Therefore, one may speculate along with Indreko that proto-Finno-Ugrians, carriers of the Combed Ware culture, lived in the area of the present-day Latvia and Lithuania. Later on, the wave of expansion of the bearers of the Boat Axe or Corded Ware culture reached the area of the previous Finno-Ugric areas in Lithuania and Latvia between 2 500…1 800 BC. The Corded Ware people are considered to be the original forebears of the Indo-European speakers (Indreko 1948). By Wiik (2000), the process of language shift from Finno-Ugric to Baltic had reached the area of Lithuania and Southern Latvia by that time (about 2 000 BC), and populations residing in this region probably acquired new language, while largely maintaining their genetic identity. Their possible mixture with proto-Slavs is suggested by a rather high frequency of haplogroup 3 in Latvians (about 40%, Zerjal et al. 1999). Indeed, according to our data (Rootsi et al. 2000), high frequency (35–50%) of haplogroup 3 is typical of Slavic populations.

Hungarian population belongs linguistically to the Ugric branch of the Finno-Ugric language speakers. Hungarians turned out to be an exception among Finno-Ugrians as for the frequency of *Tat* C: it is very low, while the distribution of their other Y-chromosomal markers is typical of an "average" Central-European population (see Table 2).

Circum-Arctic spread of Tat C

By the efforts of several groups much has been done to understand the spread of *Tat* C allele in northern Eurasia. It is clear by now that this variant of the Y chromosome is truly circum-Arctic (see the map in Fig. 6). Being by far the dominant variety of Y chromosome in this area, its spread is not restricted to any linguistically defined set of populations: it can be found among Uralic, Indo-European and Altaic-language speaking populations of the region. Thus, at least in respect of this marker, the spread of maternally and paternally inherited genes in humans differs significantly. The question is: how did such cross-linguistic spread of a certain Y-chromosomal variety come to be in the first place? The most puzzling is its circum-Arctic spread. One possible explanation for the spread of *Tat* C is a "substratum" concept: a branch of *Tat* C carriers migrated from the Eastern European Plain refugium area to north and east, so to say, following the path of the migrating big game. Alternatively, the spread of *Tat* C might have been influenced by some kind of gene that is under natural selection: its carriers had better resistance to cold climate and therefore this variant of *Tat* was simply "*hitchhiking*". There may be other speculations as well, but it is obvious that the phylogeography of the distribution of *Tat* C chromosomes deserves a great deal of attention by those wanting to understand genetic history of the Nordic people.

Conclusions

Our contribution presented above was written bearing in mind a strictly specific context: it consists of a selection of comments to one relatively short paper published more than half a century ago by Richard Indreko. While speaking about the "ideas of Indreko", we are well aware that the paper by Indreko is a review, and the ideas suggested in it belong to many archaeologists and were formulated over a considerable period of time. Moreover, it is obvious that the last 50 years have brought much new knowledge to the understanding of the prehistory of Europe. We did not, however, think it reasonable here to try to "update" Indreko by incorporating into our paper references to a wealth of new archaeological and linguistic arguments. Instead, we took his paper at face value and made an attempt to analyse a selected number of ideas presented in it from the point of view of a rapidly accumulating new genetic knowledge. Furthermore, it was certainly not our task or even an attempt to choose genetic facts specifically supporting Indreko's views: we would have been equally happy to challenge whatever he said. Somewhat surprisingly even to ourselves, we found a sound concordance between some basic suggestions made by Indreko and our current understanding of the genetic history of northern Europeans. Moreover, we were forced to admit that in some other cases where we could not reach any clear position, the problem was not in Indreko´s ideas but in the nature of our genetic evidence. At this point we would only like to add that the tools available for human population genetics will exponentially increase in the very near future, giving cause to hope that its heuristic value would soon be lifted to a level that allows to seek explanations to the questions raised by Indreko that we left unanswered.

Address:
Kristiina Tambets
Institute of Molecular and Cell Biology
University of Tartu and Estonian Biocentre
Riia 23, Tartu 510101
Ph: +372 07 375 053
E-mail: ktambets@ebc.ee

References

Aul, J. (1964) "Antropologia Estoncev". *TRÜ Toimetised* 158.

Bandelt, Hans Jürgen, Forster, P., Sykes, B. C. and Richards, M. B. (1995) "Mitochondrial Portraits of Human Populations Using Median Networks". *Genetics* 141, 2, 743–53.

Bertranpetit, Jaume, Sala, J., Calafell, F., Underhill, P. A., Moral, P. and Comas, D. (1995) "Human mitochondrial DNA variation and the origin of Basques". *Annals of Human Genetics* 59, Pt 1, 63–81.

Cann, Rebecca L., Stoneking, M. and Wilson, A. C. (1987) "Mitochondrial DNA and Human Evolution". *Nature* 325, 6099, 31–6.

Cavalli-Sforza, Luigi Luca, Menozzi, P. and Piazza, A. (1994) *The History and Geography of Human Genes*. Princeton: Princeton University Press.

Comas, David, Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E., Martinez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., Pettener, D. and Bertranpetit, J. (1998) "Trading Genes Along the Silk Road: mtDNA Sequences and the Origin of Central Asian Populations". *American Journal of Human Genetics* 63, 6, 1824–38.

Corte-Real, H. B., Macaulay, V. A., Richards, M. B., Hariti, G., Issad, M. S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J. and Sykes, B. C. (1996) "Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis". *Annals of Human Genetics* 60, Pt 4, 331–50.

Indreko, Richard (1948) "Origin and Area of Settlement of the Fenno-Ugrian Peoples". In *Science in Exile. Publication of the Scientific Quarterly "Scholar" 1*. 3–24. Heidelberg.

Jobling, Mark A., Pandya, A. and Tyler-Smith, C. (1997) "The Y Chromosome in Forensic Analysis and Paternity Testing". *Inernationalt Journal of Legal Medicine* 110, 3, 118–24.

Karafet, Tanya M., Zegura, S. L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., de Knijff, P., Wiebe, V., Griffiths, R. C., Templeton, A. R. and Hammer, M. F. (1999) "Ancestral Asian Source(s) of New World Y-Chromosome Founder Haplotypes". *American Journal of Human Genetics* 64, 3, 817–31.

Kittles, Rick A., Bergen, A. W., Urbanek, M., Virkkunen, M., Linnoila, M., Goldman, D. and Long, J. C. (1999) "Autosomal, Mitochondrial, and Y Chromosome DNA Variation in Finland: Evidence for a Male-Specific Bottleneck". *American Journal of the Physical Anthropology* 108, 4, 381–99.

Kivisild, Toomas, Bamshad, M. J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., Laos, S., Parik, J., Watkins, W. S., Dixon, M. E., Papiha, S. S., Mastana, S. S., Mir, M. R., Ferak, V. and Villems, R. (1999) "Deep Common Ancestry of Indian and Western-Eurasian Mitochondrial DNA Lineages". *Current Biology* 9, 22, 1331–1334.

Lahermo, Päivi, Sajantila, A., Sistonen, P., Lukka, M., Aula, P., Peltonen, L. and Savontaus, M. L. (1996) "The Genetic Relationship Between the Finns and the Finnish Saami (Lapps): Analysis of Nuclear DNA and mtDNA". *American Journal of Human Genetics* 58, 6, 1309–22.

Lahermo, Päivi, Savontaus, M. L., Sistonen, P., Beres, J., de Knijff, P., Aula, P. and Sajantila, A. (1999) "Y Chromosomal Polymorphisms Reveal Founding Lineages in the Finns and the Saami". *European Journal of Human Genetics* 7, 4, 447–58.

Lahr, Marta and Foley, R. (1994) "Multiple Dispersals and Modern Human Origins". *Evolutionary Anthropology* 3, 48–60.

Lutz, Sabine, Weisser, H. J., Heizmann, J. and Pollak, S. (1998) "Location and Frequency of Polymorphic Positions in the mtDNA Control Region of Individuals from Germany". *International Journal of Legal Medicine* 111, 2, 67–77.

Macaulay, Vincent A., Richards, M. B., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonné-Tamir, B., Sykes, B. and Torroni, A. (1999) "The Emerging Tree of West Eurasian mtDNAs: A Synthesis of Control-Region Sequences and RFLPs". *American Journal of Human Genetics* 64, 1, 232–49.

Mark, Karin (1956) "Eesti rahva etnilise ajaloo küsimusi paleoantropoloogia valgusel". In *Eesti rahva etnilise ajaloo küsimusi.* 191–211, Tallinn.

Metspalu, Ene, Kivisild, T., Kaldma, K., Parik, J., Reidla, M., Tambets, K. and Villems, R. (1999) "The Trans-Caucasus and the Expansion of the Caucasoid-Specific Human Mitochondrial DNA". In *Genomic Diversity: Application in Human Population Genetics.* Papiha, S., Deka, R. and Chakraborty, R., eds. 121–134, New York: Kluwer Academic / Plenum Publishers.

Nei, Marsatoshi (1987) *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Opdal, Siri Hange, Rognum, T. O., Vege, A., Stave, A. K., Dupuy, B. M. and Egeland, T. (1998) "Increased Number of Substitutions in the D-loop of Mitochondrial DNA in the Sudden Infant Death Syndrome". *Acta Paediatrica* 87, 10, 1039–44.

Orekhov, Vladimir, Poltoraus, A., Zhivotovsky, L. A., Spitsyn, V., Ivanov, P. and Yankovsky, N. (1999) "Mitochondrial DNA Sequence Diversity in Russians". *FEBS Letters* 445, 197–201.

Piercy, Romelle, Sullivan, K. M., Benson, N. and Gill, P. (1993) "The Application of Mitochondrial DNA Typing to the Study of White Caucasian Genetic Identification". *International Journal of Legal Medicine* 106, 2, 85–90.

Quintana-Murci, Lluis, Semino, O., Bandelt, H. J., Passarino, G., McElreavey, K. and Santachiara-Benerecetti, A. S. (1999) "Genetic Evidence of an Early Exit of Homo Sapiens Sapiens from Africa Tthrough Eastern Africa". *Nature Genetics* 23, 4, 437–41.

Richards, Martin, Corte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H. J. and Sykes, B. (1996) "Paleolithic and Neolithic Lineages in the European Mitochondrial Gene Pool". *American Journal of Human Genetics* 59, 1, 185–203.

Richards, Martin, Macaulay, V., Sykes, B., Pettit, P., Forster, P., Hedges, R. and Bandelt, H. J. (1997) "Reply to Cavalli-Sforza and Minch". *American Journal of Human Genetics* 61, 251–254.

Richards, Martin B., Macaulay, V. A., Bandelt, H. J. and Sykes, B. C. (1998) "Phylogeography of Mitochondrial DNA in Western Europe". *Annals of Human Genetics* 62, Pt 3, 241–60.

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

Rootsi, Siiri, Tambets, K., Adojaan, M., Orekhov, V., Khnusnutdinova, E., Yankovsky, N., Villems, R. (in press) "Archaeogenetics of Finno-Ugric Speaking Populations". In *The Roots of Peoples and Languages of Northern Eurasia IV*. K. Julku, ed. Turku.

Rousselet, F. and Mangin, P. (1998) "Mitochondrial DNA Polymorphisms: A Study of 50 French Caucasian Individuals and Application to Forensic Casework". *International Journal of Legal Medicine* 111, 6, 292–8.

Sajantila, Antti, Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M. L., Aula, P., Beckman, L., Tranebjaerg, L., Gedde-Dahl, T., Issel-Tarver, L., DiRienzo, A. and Paabo, S. (1995) "Genes and Languages in Europe: An Analysis of Mitochondrial Lineages". *Genome Research* 5, 1, 42–52.

Sajantila, Antti and Paabo, S. (1995) "Language Replacement in Scandinavia". *Nature genetics* 11, 4, 359–360.

Salas, Antonio, Comas, D., Lareu, M. V., Bertranpetit, J. and Carracedo, A. (1998) "mtDNA Analysis of the Galician Population: A Genetic Edge of European Variation". *European Journal of Human Genetics* 6, 4, 365–75.

Schurr, Theodore G., Sukernik, R. I., Starikovskaya, Y. B. and Wallace, D. C. (1999) "Mitochondrial DNA Variation in Koryaks and Itel'men: Population Replacement in the Okhotsk Sea-Bering Sea Region During the Neolithic". *American Journal of the Physical Anthropology* 108, 1, 1–39.

Setälä, Emil, N. (1926) "Johdanto". In *Suomen suku I*. Kannisto, A., ed. Helsinki: Otava.

Tambets, Kristiina, Kivisild, T., Metspalu, E., Parik, J., Kaldma, K., Laos, S., Tolk, H.-V., Gölge, M., Demirtas, H., Geberhiwot, T., Papiha, S., S., De Stefano, G., F. and Villems, R. (in press) "The Topology of the Maternal Lineages of the Anatolian and Trans-Caucasus Populations and the Peopling of the Europe: Some Preliminary Considerations". In *Molecular Genetics and Early Europe: Papers in Population Prehistory*. Cambridge: McDonald Institute for Archaeological Research Monograph Series, Cambridge University.

Torroni, Antonio, Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M. L. and Wallace, D. C. (1996) "Classification of European mtDNAs from an Analysis of Three European Populations". *Genetics* 144, 4, 1835–50.

Torroni, Antonio, Petrozzi, M., D'Urbano, L., Sellitto, D., Zeviani, M., Carrara, F., Carducci, C., Leuzzi, V., Carelli, V., Barboni, P., De Negri, A. and Scozzari, R. (1997) "Haplotype and Phylogenetic Analyses Suggest that one European-specific mtDNA Background Plays a Role

in the Expression of Leber Hereditary Optic Neuropathy by Increasing the Penetrance of the Primary Mutations 11778 and 14484". *American Journal of Human Genetics* 60, 5, 1107–21.

Tyler-Smith, Chris (1999) "Y-Chromosomal DNA Markers". In *Genomic Diversity: Application in Human Population Genetics.* Papiha, S., Deka, R. and Chakraborty, R., eds. 65–73, New York: Kluwer Academic / Plenum Publishers.

Villems, Richard, Adojaan, M., Kivisild, T., Metspalu, E., Parik, J., Pielberg, G., Rootsi, S., Tambets, K. and Tolk, H.-V. (1998) "Reconstruction of Maternal Lineages of Finno-Ugric Speaking People and Some Remarks on Their Paternal Inheritance.". In *The Roots of Peoples and Languages of Northern Eurasia I.* Wiik, K. and Julku, K., eds. 180–200, Turku.

Wiik, Kalevi (2000) "European Lingua Francas". In *The Roots of Peoples and Languages of Northern Eurasia II and III*. Ago Künnap, ed. 202–236, Tartu.

Zerjal, Tatiana, Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F. R., Schiefenhovel, W., Fretwell, N., Jobling, M. A., Harihara, S., Shimizu, K., Semjidmaa, D., Sajantila, A., Salo, P., Crawford, M. H., Ginter, E. K., Evgrafov, O. V. and Tyler-Smith, C. (1997) "Genetic Relationships of Asians and Northern Europeans, Revealed by Y-chromosomal DNA Analysis". *American Journal of Human Genetics* 60, 5, 1174–83.

Zerjal, Tatiana, Pandya, A., Santos, F. R., Adhikari, R., Tarazona, E., Kayser, M., Evgafov, O., Singh, L., Thangaray, K., Destro-Bisol, G., Thomas, M. G., Qamar, R., Mehdi, S. Q., Rosser, Z. H., Hurles, M. E., Jobling, M. A. and Tyler-Smith, C. (1999) "The Use of Y-Chromosomal DNA Variation to Investigate Population History: Recent Male Spread in Asia and Europe". In *Genomic Diversity: Applications in Human Population Genetics.* Papiha, S., Deka, R. and Chakraborty, R., eds. 91–101, New York: Kluwer Academic / Plenum Publishers.