HISTORICA FENNO-UGRICA

The Roots of Peoples and Languages of Northern Eurasia

I

Turku 30.5.-1.6.1997

Edited by Kyosti Julku & Kalevi Wiik

Editorial Assistant
Merja Äärela

Societas Historiae Fenno-Ugricae
Turku 1998

Richard Villems, Maarja Adojaan, Toornas Kivisild, Ene Metspalu, Jüri Parik, Gerli Pielberg, Siiri Rootsi, Kristiina Tambets and Helle-Viivi Tolk

Reconstruction of Maternal lineages of Finno-Ugric speaking people and some remarks on their Paternal inheritance

Summary

Analysis of maternal and paternal lineages of Estonians and other North European Finno-Ugric speaking people reveal their close genetic relatedness. This conclusion is also true for Saami, although genetic drift perhaps does not allow to see it at the first glance. Maternal lineages of Finno-Ugrians are predominantly a subset of these ,found all over Europe. Among paternal lineages an intriguing link with some Siberian populations was suggested recently by others. Our analysis supports an explanation that a particular mutation under discussion arose in proto-Finno-Ugrians long ago and only later spread eastward.

Introduction

Anthropology is traditionally identified as a discipline, encompassing archaeology, physical anthropology and cultural & social anthropology. Somewhere very close stands linguistics: so close that it is only essential for anthropologists to know, at least in most general terms, what is going on in linguistics and *vice versa.*

For a variety of objective reasons, the question of the roots of Finno-Ugrians is a perfect example where linguistics, physical anthropology and archaeology must meet. It is not our task here to elaborate this statement at any length. Instead, we would like to point to some specific problems, more directly related to our work in Tartu.

Homo sapiens sapiens, modern humans, are thought to have originated some 150 000-200 000 years before present (YBP) in Africa and to have spread over Eurasia starting some 100 000 YBP *(Cann* et al. 1987, *Vigilant* et al. 1991, *Cavalli-Sforza,* 1997) At least that is what the "Recent Out of Africa" current standard model tells us. Details of the early events of this process are poorly understood and "Weak Garden of Eden" versions, which allow, e.g., several recent out of Africa migrational waves (including even occasional "back to Africa" migrations of some populations), are currently under discussion *(Harpending* et al. 1993, *Hammer* et al. 1997). One of the critical aspects here is the dating of divergence of the main races (e.g. *Nei* & *Roychoudhuri,* 1993). In this context we are specifically interested in early branching of Caucasoids and Mongoloids and in the radiation of Caucasoids. Using from the point of population genetics not entirely justified terms (Caucasoids, Indo-Europeans etc), it is very much needed to get by far better insight into radiation of Caucasoids to Indo-European, Afro-Asiatic, Finno-Ugric, Caucasian etc. (language speaking) populations.

We need to be ready for hypotheses, which are truly novel and are not bound to assumptions derived from either linguistic or genetic or archaeological data and arguments alone.

The first problem we would like to address here is the question of genetic relatedness and demographic history of Finno-Ugric languages speaking people. The second is the place of Finno-Ugrians among Caucasoids in general and their genetic relatedness to Indo-European speakers in particular. Note that where speaking about Finno-Ugrians, we refer to language and not to genetic affinity. Also, unless specified otherwise, where speaking about genetic relatedness we specifically keep in mind maternal and paternal inheritance. This is a somewhat narrow angle, but, as will be explained below, it has some inherent advantages.

Human genetic polymorphism is an empirical observation recognised obviously very long ago, perhaps as long as the history of the species. And in a way even earlier, since kinship recognition is a well established biological phenomenon among mammals. As a science, it found its beginning in the paper by *Hirszfeld* & *Hirzfeld* from 1919, revealing differences in the distribution of certain blood groups among races. Since then, several generations of scientists contributed to the worldwide collection of data on frequencies of "classical" genetic markers among populations. The latest impressive summary of this enormous work was published only recently by *Cavalli-Sforza, Menozzi* and *Piazza* (1994).

Nevertheless, only approximately 10 years ago a new period has started and this is already based on cloning, sequencing and amplification of DNA – to name the key technical novelties in the order of their invention. These methods, as we have all reasons to believe, allow to turn a new page in physical anthropology. Being still at the beginning of the road, it is already clear that the first decade was, by and large, the decade of mitochondrial DNA (below mtDNA) polymorphisms. Only a few years ago a definite breakthrough has occurred in exploring Y chromosome for anthropological search – potentially at least as informative as exploiting mtDNA *(Jobling* & *Tyler-Smith* 1995). And it does not need much imagination to foresee that billions of nucleotides in autosomal chromosomes are a source of practically unlimited new insights into understanding all aspects of human biology for decades ahead.

Despite of the potential wealth of information buried into autosomal DNA, specific features of the mitochondrial DNA and the Y chromosome make them superior to the autosomal genes in an important context, where direct maternal and paternal genetic inheritance is investigated. It is because of two reasons. Firstly, for all we know, human mtDNA is inherited maternally, while the Y chromosomal DNA is passed from fathers to sons only (here, we are not going into details like pseudoautosomal part of this chromosome). Secondly, both of these DNAs differ from that in autosomal chromosomes in an additional important aspect: they do not recombine. As a consequence, while our autosomal and X-chromosomal genes have a very large number of ancestors, all mtDNA genomes and Y chromosomes currently in circulation among contemporary mankind, are direct descendants from a single woman and a single man. Or not necessarily even *Homo sapiens sapiens:* the Most Recent Common Ancestor for these particular genomes can as well be, at least theoretically, a predecessor of our species. Moreover, it is unlikely that our "mitochondria1 Eve" and the "Y-chromosomal Adam" lived even approximately at the same time. Quite surprisingly, the mean time to the ancestral Y chromosome was estimated to be approximately 190 000 YBP (*Hammer* 1995) and although the other similar paper *(Whitfield* et al. 1995) ended up with a much shorter time depth (37 000-49 000 YBP), the latter authors' data were reinterpreted with a result closer (–160 000 YBP) to that obtained by *Hammer (Tavaré* et al. 1997). Diversity (polymorphism) of the mtDNA and the Y chromosome that we exploit, has been, therefore, generated during the last 5000-8000 generations or so. Hopefully mainly stochastically.

There is one more introductory remark of a general nature that deserves mentioning. Namely,

it is increasingly more obvious - and that is a lesson already learned by the community - that obtaining new empirical information is but a half of the problem. The other half is a much bigger one, and that is how to interpret the collected data, be them DNA sequences, microsatellite frequencies, transpositional events like Alu insertions, or else. There is nothing novel in reminding that there are numerous pitfalls even where using "established" phylogenetic methods. Besides that, in many cases "a brute computational force" is out of question simply because even the best available computers reach easily the margin of practical uncomputability where dealing with complex phylogenetic trees. Furthermore, the treebuilding allows at best to reconstruct histories of genetic lineages: linking them with the peopling of the world by modern *Homo sapiens,* with ancient and recent migrations and population dynamics in general, is even more complicated. This is why any thorough population genetic analysis should take into account and learn, as much as it is possible, from history, archaeology in particular, social anthropology and linguistics (see. e.g. *J. Diamond,* 1997).

This paper, based on a talk given by one of us (R. V.) in the symposium hold in Turku, May 29.-31. 1997, has the following structure. First, we present our data on mtDNA D-loop sequence divergence among Estonian population and we also present information about RFLP data on the same population. Secondly, we show the median network tree of D-loop sequence variants of Estonians, as well as a median network tree(s) of combined datasets. Here, two different phylogenetic reconstructions are presented: Estonian data at the background of other Finno-Ugric populations, as well as Estonian data at the background of selected representative data about the European Indo-European speaking populations and Caucasoids in general. This chapter ends with conclusions drawn from our lineage reconstructions.

In the second part of the paper we bring our new results on the frequencies of the Y chromosome-specific markers: microsatellites. insertions, deletions and on one specific point mutation. We discuss these results, as far as it is possible at present, in the light of recent worldwide data on the Y chromosome markers, paying specific attention to "the Asian 50f2/C" and "the T versus C allele" (*Jobling* et al.1996, *Zerjal* et al. 1997).

What we inherited from our mitochondrial Eve ?

1. Estonian Mothers

An absolutely strict maternal inheritance of the mitochondrial DNA in humans is still debated *(Ankel-Simons & Cummins,* 1996). However, even if an occasional sperm mtDNA genome does find its way to the fertilized oocyte (though there is no experimental evidence for that), one should keep in mind that there are perhaps as many as 100 000 mtDNA molecules in an unfertilized oocyte and less than 10 in a sperm. Therefore, for all practical purposes one may ignore the problem anyway. Major demographic events like bottlenecks and founder effects in new expansions, drift in general and in isolated Palaeolithic populations in particular - all of them may have and certainly have much larger and not always well understood effects, when we analyse particular human populations *(Wakeley & Hey,* 1997; *Relethford & Harpending,* 1995)

Using various approaches we analysed mtDNA of more than 1000 Estonians out of our "gene bank" of several tens of thousands. About 120 of them were examined by us "in full" - i.e. by combining D-loop hypervariable region 1 (HVR 1) direct sequencing with restriction fragment length polymorphism study for the same individuals. Keeping in mind possible geographic variations, the sample was combined from individuals from Virumaa, Saaremaa, Põlvamaa. Pärnumaa, Viljandimaa and Tartumaa. i.e. covering Estonia in both geographic directions. It can be immediately told that no gross differences

in this respect were found: typical haplogroups and even lineages can be found in all mentioned locations without any clear clines.

Like elsewhere in Europe, Estonian maternal lineages are highly variable: out of 100 HVR 1 sequences we found 61 variants (haplotypes), shown in Fig. 1. To analyse and to present the results graphically, we employ median network analysis of the DNA sequences *(Bandelt* et al. 1995) in combination with the RFLP data. For the latter, we use defined by *Torroni* et al. (1994, 1996) haplogrouping, since it allows rather clearly not only to organise the data, but also to distinguish European (Caucasoid) mtDNA varieties from "genuine" Asian and African haplotypes. None of the borders are absolute and even the most widespread in Europe "Cambridge Reference Sequence"-type mtDNA (CRS; *Anderson* et al. 1981) variant was at least once found in sub-Saharan Africans as well as elsewhere in non-Caucasoids. However, in general the European-specific haplogroups H, I, J, K, T, U, V, W, and X are much more frequently found in Caucasoids than in Mongoloids or Africans (Torroni et al. 1996). And *vice versa* - haplogroups typical in Africans and Mongoloids like L, A, M, are very seldom seen in Europens. One must be specifically careful where speaking about Caucasoids or using terms as "Indo-European (speaking) populations": already what little we know about India at present, tells that one can find European, Asian, African and perhaps hitherto undefined mtDNA variants in the more than billion inhabitants of this subcontinent. Table 1 compares frequencies of Estonian mtDNA haplogroups with some other European populations. The choice of these other populations is limited for the reason that we wanted to be "on a safe side" and used for comparison only the data recently published by Torroni et al. (1996): for other populations only partial data can be indirectly derived from the HVR I sequence similarities, but not directly from the RFLP data.

So where are the differences? There are some, but not reliably distributed either geographically or by language phyla.

Table 1. *Distribution of human mitochondrial DNA haplogroups in some European populations. Haplogroup nomenclature according to Torroni et al. (1996).*

Haplogroup	Estonians	Finns (as % from sample)	Tuscanians	Swedes
H	51.0	40.8	41.7	40.5
I	1.5	2.0	4.2	0
J	7.0	14.3	14.6	2.7
K	3.0	4.1	6.3	13.5
M	0	2.0	0	0
T	6.0	6.1	10.4	21.6
U	21.0	16.3	10.4	16.2
V	2.0	4.1	0	5.4
W	2.5	4.1	2.1	0
X	1.0	4.1	2.1	0
others	5.0	2.0	2.1	0

Notes:
a) haplogroup M is considered as "Asian" (Torroni et al. 1994, 1996)
b) sample sizes are: Estonians 200; Finns 49; Tuscanians 48; Swedes 37

As mentioned above, in analysing sequence data we prefer median network analysis (*Bandelt* et al. 1995). Median networking is in principle parsimonious tree analysis, but instead of constructing a number (sometimes a very large number) of different trees, median network tree depicts a number of most parsimonious trees simultaneously. There can easily be hundreds of most parsimonious trees visualised and easily followed in a single picture, whereas ambiguities are also clearly shown as "reticulations" – parallel routs to the nods of the graph.

"Reduced" median network of Estonian mtDNA haplotypes is depicted in Fig. 1. This network was constructed without using RFLP data. Different geographic locations of individual haplotypes are indicated. Bars without any indication are Estonian sequences described by others (*Sajantila* et al. 1995). The figure shows typical for Europe divergent network of mtDNA haplotypes, grouped into haplogroups. Here we use the numeration proposed by *Richards* et al. (1996). Three-digit numbers refer to the mitochondrial genome D-loop hypervariable segment I (HVS I) sequence numbers according to Cambridge Reference Sequence (*Anderson* et al. 1981). For simplicity, the first two digits are omitted, so that, e.g., number 126 defines position 16126 in the mtDNA genome.

Adding RFLP data to the sequencing data, one ends up with seemingly more complex network (not shown here). Nevertheless, this "combined" network allows greater details and in fact simplifies the overall result by reducing ambiguities. As an example, haplogroup 2 can be clearly separated into two distinct haplogroups J and T (Fig. 2). Also, the array of lineages bearing substitution at np 16223 are grouped into separate lineages etc. etc. It is important to note that this is not merely a problem of better resolution. For example, on the basis of the topology of haplogroup 2 of the first network one could suggest that lineages leading to the branches 2A and 2B have a common ancestor, differing from the central nod (here: from the CRS) by a single mutation at np 16126 (Fig. 1). Fig. 2, however, shows that the seemingly common for 2A (= J) and 2B (= T) root is in fact split by several mutations: gains and losses of novel compared to CRS restriction enzyme cutting sites. Therefore, the two maternal lineages of Estonians starting with the mutation at np 16126, are not necessarily related. Moreover, though J and T are European-specific mtDNA lineages, their ancestral form(s) have not been found (at least yet) in Europe. This split has likely already happened in Paleolithic. Of course, "unrelated is a relative term: the very idea of our phylogenetic tree-building originates from a basic assumption that all extant human beings have a single common "mitochondrial" ancestor.

In sum, Estonian maternal lineages are an example of a collection of typical European mtDNA lineages and their genetic distance from those found in most of the other European populations is small indeed (see also *Sajantila* et a1.1995). So far we have found only a few Estonian lineages out of approximately 170, which have no obvious match with described by Torroni et al. (1996) Caucasoid mtDNA haplogroups. Table 2 documents aforementioned lack of significant genetic differences between maternal lineages in Estonians from various geographic locations, as well as between those and sampled in North Germany.

Pairwise mismatch distribution of gene sequence differences (Table 2, Figure 3) is a method thought to reflect in a not yet fully understood way the demographic history of populations (*Rogers & Harpending* 1992, *Harpending & Rogers* 1995). For Estonians, the histogram of the distribution (Fig. 3) is what one can foresee for a population which underwent an early exponential growth phase and after that never faced severe bottlenecks. This type of curve is typical for most of European populations and is an additional circumstantial evidence that the coalescence (and the divergence time) of the maternal lineages of Estonians date by and large back to the Paleolithic, likewise it was shown now for European populations in general (*Richards* et al. 1996, *Torroni* et al. 1996).

***Table** 2. Mean mismatch (diagonal) and intermatch (below diagonal) distribution of pairvise sequence differences in human mitochondrial DNA D-loop hypervariable region I in different populations, and their genetic distances (above diagonal).*

	Saaremaa	Põlva	Viljandi	Estonia[1]	N.Germany[2]
Saaremaa	**5.40**	–0.06	0.07	–0.05	0.04
Põlva	4.74	**4.19**	0.06	–0.07	–0.03
Viljandi	4.97	4.35	**4.38**	0.004	0.11
Estonia	4.91	4.28	4.37	**4.52**	0.03
N.Germany	4.67	3.99	4.14	4.21	**3.85**

[1] – *Estonian data from this laboratory,* n = *99*
[2] – *N. German data from Richards et* al. (1996), *n = 107*

The number of mtDNA haplogroups present in any population is largely a question of defining the former and therefore arbitrary. We are at present satisfied with the classification of European mtDNA sequences worked out by *Torroni* et al. (1994, 1996). Its further refinement is certainly possible (e.g. subdivision of already too bulky H, subdivision in haplogroup U), but it would be most useful if the "mtDNA community" would agree to follow this nomenclature, together with HVR1 – based lineage analysis. The question about the total number of mtDNA D-loop HVRl sequence haplotypes possibly present in Estonian population remains. Numerical simulation of Estonian data showed us that we are still far from midpoint of the possible total number. On the other hand, any novel haplotype would very likely belong to haplogroups, already found in Estonians. Therefore, it is at present scientifically more interesting to compare Estonian data set with similar investigations carried out in other populations, first of all in Finno-Ugric speakers.

2. The Saami Dilemma

In the context of this topic it is specifically interesting to discuss Saami data. Sequences used by us for median network reconstructions are those obtained by ourselves or published by others (*Sajantila* etal. 1995, *Lahermo* etal. 1996). We note that genetic distance estimates show very convincingly that in Europe, Saami (Lapps) are clear "genetic outsiders": their distance from others is much larger than typical genetic distances between European populations. This observation holds true both for classical genetic markers(e.g. *Cavalli-Sforza* etal. 1994) as well as for mtDNA D-loop sequence analysis (*Lahermo* et al. 1996, *Sajantila* et al. 1995). For the latter, it was found that genetic distance between Saami and any other European population studied, is very much longer than those in between other European populations, including Finns (*Sajantila* et al. 1995). This observation, supported by authors' interpretation of mtDNA phylogenetic tree, constructed by using the neighbor-joining PHYLIP program package, led to the conclusion that there is *"a considerable difference between the genetic background of the Finns and that of the Finnish Saami"* and that *"the Saami are not closely related to their linguistic and geographic neighbors"* (*Lahermo* et al. 1996).

Coming back to that stressed already above, we want to show here that these conclusions are ambiguous, at least as far as we are discussing maternal inheritance of Saami in relation to other European Finnish language group populations and, moreover, to the European mtDNA lineages in general. Let us formulate first the problem more precisely. The question we are interested in is as follows: are the Saami maternal lineages a subset of the general European network of the mtDNA lineages, or do they belong to some other,

either known (Asian or else) or hitherto unidentified lineages (haplogroups)?

The answer is quite clear: the majority of Saami mtDNAs display the same basic European set of mtDNA varieties, to which belong also Estonian, Finnish, Mari, Moksha, German, UK, Iberian etc. lineages. Furthermore, Figure 4 possibly reveals the reason of former misunderstandings: the variety of Saami lineages sampled so far is highly reduced compared to that characteristic for the other Finno-Ugric speaking populations: the majority of individuals belong to just a few haplotypes. Follow, e.g. the so-called "Saami motif' (*A. Sajantila* et al. 1995) of substitutions at nucleotide positions (nps) 16144, 16189 and 16270 (Fig. 4). It is evident that this motif is but a branch of a haplogroup, common for Europeans: it stands only one substitutional step away from a lineage shared by both the Finno-Ugric and Indo-European speakers. The other predominant Saami cluster is only one substitution away from the CRS and likely belongs, therefore, to haplogroup I, most common for Europeans (*Richards* et al. 1996). However, using another nomenclature (Torroni et al. 1996), this Saami cluster does not belong to haplogroup H, but to a minor European-specific haplogroup V (*A. Torroni* et al., in preparation). The three Saami haplotypes, which all share more common for Asia and Africa substitution at np 16223 (Fig. 4) need further refinement, e.g. by characteristic RFLP analysis, but so far it can be indicated that several "European" haplogroups (I, W & X) do possess this mutation in their D-loop sequence, whereas specific for Asians and Africans haplogroups have been sampled in Europe at a very low frequency (Richards et al. 1996, Torroni et al. 1996).

We feel that the visual impression what one gets from Fig. 4 is rather clear: both the Baltic and Volga Finno-Ugric populations (for the latter, one should keep in mind much smaller sample) are spread more or less uniformly all over the network. It is different for Saami: they are predominantly concentrated into a few clusters, **but these Saami clusters lie inside the network of maternal lineages, shared by all so far studied Finno-Ugric as well as Indo-European populations.** Moreover, we wish to stress here that this basic network is also largely shared by Afro-Asiatic language speaking populations of Algeria and Middle East, as well as Turks, who linguistically belong to Altaic people. Therefore, it is justified to say that the network is a general network of Caucasoids. Of course, here we are already crossing the margin, since we do not have enough data on many Caucasoid populations. That little what is known about India where the number of Indo-European speakers is larger than that in Europe, shows already at present that among them, "non-Indo-European" mtDNA haplogroups can be rather common (*Passarino* et al. 1996a & 1996b).

We hope that the data and analysis presented above allows to re-evaluate the problem and to conclude that maternal lineages of Finno-Ugric speaking European populations are closely related to each other and none of the so far studied populations has an apparently different genetic origin, as far as their mtDNA is concerned. Earlier discussions, in particular those concerning the genetic background of Saami, addressed in fact a different question, but were over-interpreted in terms of genetic history of Saami maternal inheritance. There is some analogy here with Basques: there, also, it was found that their maternal lineages are of the same common Iberian and European stock (*Richards* et al. 1996), although their genetic distance from any other European population, calculated on the comparative frequencies of a variety of genetic markers, is much longer than the average inside Europe (e.g. *Cavalli-Sforza* et al. 1994).

3. The place of the Finno-Ugric mothers among the mothers of Caucasoids and some general concluding remarks on this issue

Language replacement hypothesis of Finns *(Sa-*jantila & *Pääbo,* 1995) does not seem in this light any longer necessary. Moreover, we note that

while the replacement model was an interesting hypothesis to explain the Finnish/Saami genetic difference dilemma, it never, as it seems to us, helped much to understand how, then, Estonians, Mordvas, Mokshas – geographically dispersed in eastern Europe Finno-Ugric speakers – acquired their present language. Since Finns are genetically close to them, one should apply similar replacement hypothesis. The question arises and no easy answer is at hand – from where?

So what about the other way around: language replacement in northern Europe, coinciding with coming of agriculture, as hypothesised by *K. Wiik* (*Wiik*, this book)? In some general way, the mtDNA data fully support the idea of the Paleolithic continuity of European maternal lineages, irrespective of whether we speak about "Finno-Ugrians" or "Indo-Europeans". This conclusion was rather forcefully drawn in *Richards* et al. 1996, and our analysis of Estonians and others shows also that there is very little, if any, difference in maternal lineages from Cornwall to Volga. On the other hand, it is important to bear in mind that classical genetic markers – and specifically the pattern of the principal coordinate analysis – do reveal gradients. What should perhaps be stressed here, is that the first principal component, exhibiting a SE–NW gradient, covers but 28 % of the markers used (*Cavalli-Sforza* et al. 1994). Whether 28 % is a strong support, is another question. To our understanding it deserves due attention. However, the important aspect is the dating of this gradient. And here, we feel, all ends are still rather loose, so that any strong version of the *Neolithic demic diffusion* should be treated with caution.

Note that the second principal component, supported by 22 % of genetic variation of classical polymorphisms, displays a SW–NE gradient. In one of his latest papers, L. L. Cavalli-Sforza clearly links this gradient with the migration of a population speaking a Uralic family language to the northeast of Europe (*Cavalli-Sforza*, 1997).

As for the specific question of the possible switch from a proto-Finno-Ugric to a proto-Indo-European language in northern Europe some 5000 YBP, then very little can be added from the point of view of maternal lineages. And mainly because important in this context data are largely missing yet. We lack data about Poles. Sweden is to some, but still too limited extent examined (*Torroni* etal. 1996, *Sajantila* et al. 1996). Then, again, next to nothing is known about Norway. Northern Germany is covered (*Richards* et al. 1996), but also mainly in its western part; Lithuania is again not covered, etc. etc. Not to add that the last war hit hard and caused large demographic movements specifically of the populations living in the southern coast of the Baltic.

It seems to us that one lesson of possibly heuristic value is already learned: a detailed mtDNA lineage analysis rather than tabulating of genetic distances may lead to establishing of datasets, suitable for analysis of subtle relative genetic affinities between (contemporary) northern European Finno-Ugric speakers vs. Indo-European speaking northern Europeans ***vs.*** southern Europeans.

The Paternal angle

It is obvious that the Y chromosome can be at least as valuable and informative for reconstructing biological history of populations as the deciphering of mitochondria1 lineages turned out to be. Only a few years ago the number of useful polymorphic sites in this chromosome was still very limited indeed *(Spurdle & Jenkins* 1992). Now the situation has changed to better (*Jobling & Tyler-Smith* 1995) and a real boom of research has started. Because of this very limited time period, not much is done yet, even less is published. Therefore, we are also limited in our discussion.

Among those few "new wave" Y chromosome papers, published within the last year, two are of a specific interest for the Finno-Ugric roots. The first (Jobling et al. 1996) maps world-wide distribution of a specific deletion mutant of Y chromosome, designated as locus 50f2/C deletion (below: either Δ50f2/C or *DYS7C*), whereas

the other paper (*Zerjal* et al. 1997) describes the same for a point mutation in a single copy locus RBF5, resulting in T → C transition. We note that the list of the authors of these two papers largely overlap. The papers are intriguing and highly relevant to Finno-Ugric paternal inheritance because both these mutations (the deletion as well as the C allele) are not only very common for Finns, Estonians and Saami, but at the same time virtually absent west from eastern Scandinavia. And – to make it even more intriguing – these mutations are frequent in some Siberian populations (*Zerjal* et al. 1997).

The two mutations can not, in principle, cover an identical chromosomal area since one of them is a deletion. They do coincide in a specific way, allowing to suggest that people having the C allele are a subset of those, who have the deletion (*Zerjal* et al. 1997). Haplotype analysis of various populations allowed to conclude that *DYS7C* has originated independently many times at different haplotype backgrounds (*Jobling* et al. 1996). It is relatively common in eastern Asia and because of its ~10 % frequency among Han, the number of its carriers already in China alone may well be over 50 million people. Our own data show that its frequency among Russians of different geographic origin is about 18%, meaning that there can be additional 10 million Russians and Ukrainians with Δ50f2/C.

It is well known that point mutations are rare in Y chromosome (see, e.g. *Dorit* et al. 1995). Zerjal et. al. (1997) suggest that the T → C transition under discussion arose only once in the history of mankind. Furthermore, the authors also suggest that it likely had occurred first in Asia, since the background mutational event – the deletion – although arisen many times independently, seems to be strongly Asian-specific (*Jobling* et al. 1996). If indeed so, then there must be a direct genetic link between all humans with the C allele: a conclusion inevitably derived from the fact that the Y chromosome is a single haploid locus (note that although in strict sense it is not correct to speak about different loci in Y chromosome, nevertheless in everyday life people often do so). The question, then, is: when did the C allele originate? If it a is very ancient event, roughly coinciding with the branching of Mongoloids some 59 000–118 000 YBP *(Ballinger* et al. 1992), then its very specific spread can be ascribed to random drift in isolated Paleolithic populations or by other, basically stochastic demographic processes. If, on the other hand, it is indeed recent, then one should take it into any of the models, trying to explain European Finno-Ugric "roots".

Another intriguing aspect of the spread of the C allele is its geography in the East. While the Δ50f2/C can be found at medium to high frequencies in a wide geographic area, including Australia and Papua New Guinea, China and Mongolia, certain areas of Siberia like Altai and Yakutia, the C allele seems to be much less spread. First, it is not found in Australia and Papua New Guinea. Furthermore, it was so far not sampled in China and Altai, and even in most of Mongolia (*Zerjal* et al. 1997). In fact, its spread in Asia is very limited – if not geographically, since Yakutia and Buryatia cover together an area, comparable to Western Europe, then certainly in terms of populations: the C allele is found practically only among Buryats (~58 %) and Yakut (~86 %). The size of the both populations is approximately 350 000, so that taking the number of males and females equal, it leaves us approximately with a quarter of million of the C allele carriers in these two populations, taken together.

Coming back to the all-important question of dating the transition, here the authors' calculation places it ~2000-4000 YBP: a very recent event indeed. Accepting the possibility of even a substantial error in these estimates, it is nevertheless a recent date in the time scale of the peopling of Eurasia. How did the authors estimate these dates? It appears that the calculations are based on Y-chromosomal microsatellite polymorphism data, assuming that the T g C substitution occurred only once and that at this zero time, microsatellite variance at the background of the C allele was also zero. The time estimates given above were obtained, thereafter,

by extrapolating the mutation rates of autosomal tetranucleotiderepeats ($\sim 2 \times 10^{-3}$) for Y-chromosomal repeals, and using the data of the Y-chromosomal tetranucleotide variance.

Therefore, one of the critical points here is the microsatellite variance. The second is the mutation rate, estimated as an average value for a large number of autosomal repeats, hut applied here for a limited set of Y-chromosomal repeats. As for the first, the authors determined the microsatellite (incl. 6 tetranucleotide repeats) polymorphism background of 60 C-allele chromosomes and found that their variance is indeed very much lower than in Y chromosomes on average (*Zerjal* et.al. 1997).

We in our work have been interested in two problems. Concentrating basically on 2 populations, Estonians and Russians, we first asked the question about the overlap of the T g C transition with the Δ50f2/C. because, as indicated above, that number varies very significantly among populations. Secondly, we also determined the microsatellite variability for the T- and C-allele chromosomes among Estonian and Russian populations.

We found that among Estonians (n = 140). 73 % of the Δ50f2/C chromosomes have the C allele. The figure for Russians is not much different. Meanwhile, the Δ50f2/C chromosomes vary considerably: in a sample of 35 chromosomes, as many as 27 combined haplotypes can be found using only 6 polymorphic markers. In fact, they have been homologous only in regards of a recent *Alu*I repeat insertion: all were YAP-negative and only a tew combined haplotypes were found more than once.

Looking to the data of Zerjal et al. (1997). one can find that in their sample all C allele carriers are homologous for the *DYF371* repeat polymorphism and belong to class 198, 201. And that irrespective to whether the individuals are Yakut. Buryats, Finns or others. Among Estonian C-allele chromosomes, while 77 % belong to this class, both 195, 201, 210 and 195, 201, 205, 207 classes were found at frequencies of about 10 % (Table 3).

Here, the precise distribution of the frequencies of *DYF371* repeat lengths is not very important factor: the message is that in the Estonian sample there is a significant heterogeneity already in this Y chromosomal marker among the *DYS7C* C allele chromosomes.

Furthermore, we found a number of other tetranucleotide repeat variations (new combination microsatellite haplotypes compared to Zerjal el al. 1997) as well as a more "smooth" distribution of them in virtually all *DYS7C*/microsatellite combination haplogroups (Table 3). Since these details are very important for the estimation of the time of the origin of the C allele, we comment them.

For example, while in Zerjal et al.(1997). 98,5 % of C allele chromosomes belong to 190 nucleotide (nt) repeat length class of *DYS19,* then among Estonians 33 % have the length of 194 nt (Table 3). One can agree that the founding C-allele chromosome had 190 nt long *DYS19,* but our data suggest a much larger share of an additional mutation step in this tetranucleotide repeat, meaning again a need for an earlier dating of the T gC transition. Also, please note that the slight heterogeneity in the other dalaset is caused by one Finnish sample (*Zerjal* et al. 1997).

For *DYS389a,* the predominant C allele-specific size class in Zerjal et al.(1997) paper is 251 nt (92 %), whilst in Estonian sample 59 % chromosomes belong to this size class and 35 % have repeat length 255 nt. In addition, there are other *DYS389a* variants in Estonians (Table 3). Accepting the single origin hypothesis, we suggest that either sampling effect or random drift explains this indeed very narrow distribution of allelic frequencies of *DYS389a,* observed by Zerjal et al.(1997).

The same general picture holds for *DYS390*: although in both studies the predominant repeal length was 211 nt, the Estonian sample has a significant proportion of C allele chromosomes with repeat length 219 nt (Table 3). Moreover, here again the heterogeneity in the other study is solely caused by the Finnish individuals in their sample.

Table 3. *Length distribution of some Y chromosomal microsatellites at the background of the C allele of DYS7C*

combination haplotype (*DYS7C* and...)	Frequency, % in Estonians	by Zerjal et al.[1]
DYF371		
198,201	77	100
194,204	0	0
194,201,210	10	0
194,201,204,207	13	0
DYS389a		
247	3	0
251	59	8
255	35	92
259	3	0
DYS390		
207	0	2
211	73	78
215	17	18
219	10	2
DYS19		
190	67	98
194	33	2

[1] ref: Zerjal et al. (1997)

As for *DYS391*, it is true that the number of its variants is low: in case of the C allele chromosomes it is (almost) limited to lengths classes 283 and 287 nt, present at frequencies 31 % and 69 % in our study and ~43 % and ~57 % in Zerjal et al.(1997), respectively. However, in an Estonian random sample of 100 Y chromosomes, these frequencies are 45 % and 55 %, respectively. Therefore, it seems that the variance of *DYS391* at the background of the C-allele chromosomes, starting possibly from 100% of either one of two *DYS391* variants, may already be close to equilibrium. It has not likely happened within ~3000 years - unless the rate of mutation of this particular tetranucleotide is much faster than anticipated. We may add that the worldwide study of the distribution of the *DYS391* repeat lengths revealed that the size class 283 nt is the predominant one, varying in most of the populations between ~53 % to 86 % *(Deka* et al. 1996).

How should one interpret all these findings'? It would be premature and perhaps also unnecessary to suggest multiple origins of the C allele. After all, like Zerjal et al.(1997), we also saw that the C allele is a subset of A50f2/C-type chromosomes. What we want to question, is the timing of the occurrence of this T —>C transition. Our data show that although the C allele chromosomes have less variable microsatellite diversity background, compared to that for Estonians in general, this is still much wider than it appears from the results published by the other group. As indicated above, microsatellite divergence is at present the critical and in fact the only source to estimate the

time of the origin of the C allele. Our results suggest significantly earlier dates.

What about the place of origin? Jobling et al. (1996) brought clear evidence for multiple origins of A50f2/C. Since this deletion was not found in Africa, they suggested that all this type deletion events occurred in Asia. We have, in fact, an additional problem to solve in case of *DYS7C*, found in Estonians. The same is apparently true for North European Finno-Ugric speakers in general. Namely: is this deletion homogeneous or do we have several similar, but independently arisen (in Asia?) polymorphisms? Rigorous answer needs a careful additional investigation of this problem. However, let us assume here for simplicity that we have in northeastern Europe only a single prototype of *DYS7C*. We can also assume that the deletion mutant had, *in statu nascendi*, the T allele. One does not need to take the latter for granted, but in our Estonian sample we indeed saw at the T allele background a larger variability of the Y chromosomal microsatellites, than we observed in the C allele chromosomes (Table 4).

These observations and arguments, taken together, allow us to question the Mongolian origin of the C allele, suggested by Zerjal et al. (1997). Instead, we suggest that this polymorphism has its origin among proto-Finno-Ugric populations of northeastern and eastern Europe, from where it, considerably later, found its way to some selected Siberian populations. This explanation allows to understand why the incidence of the C allele is high among Finno-Ugric populations and, what is more important, why it is significantly more diverse there, than among Yakut and Buryats.

Meanwhile, we note that "East-Asian affinity" of the Estonian Y-chromosomal microsatellite length variability can be suggested looking e.g. at the frequency of various *DYS19* repeats. While northern European populations have a typical unimodal distribution of *DYS19* with predominant lengths of 190 bp, Estonians have a significant proportion of 198 and 202 bp chromosomes, characteristic for East Asians as well as to certain Africans but not at all for northern Europeans (Table 5). However, this difference can be

Table 4. *Some examples of the distribution of DYS7C/microsatellite combination haplotypes at the background of C and T alleles, compared to the distribution of these microsatellites in other Estonian Y chromosomes*

repeat type & length (bp)	Estonian Y chromosomes C allele	T allele	all others
DYS390			
207	0	15.4	7.9
211	73	15.4	28.6
215	16.7	23.1	19.1
219	10	38.3	36.5
223	0	7.8	7.9
DYS389			
364	6.2	16.7	16.9
368	40.7	33.3	32.2
372	50.0	33.3	40.0
376	3.1	16.7	5.1
380	0	0	6.8

***Table** 5. Length variation of DYS19 in Estonians and in some European and Asian populations*

DYS19 repeat ...	Population		
length (bp)	Estonians[1]	N.-Europeans[2]	Asians[3]
186	3	9	4
190	35	57	16
194	26	26	39
198	24	6	32
202	12	2	9

[1] - our data, n = 174
[2] - composite of German and British populations (Hammer et al. 1997)
[3] - composite of Korean and Japanese (Aomoris) populations (from Hammer et al. 1997)

explained by drift as well. In this context it was specifically interesting to investigate the frequency and haplotype of the Y chromosomal Alu insertional polymorphism (YAP element, *DYS287)* in Estonians, described first by Hammer (1994). This insertion, designated as YAP^+, has the highest frequency in East Asians like Koreans and Japanese (Hammer et al. 1997). We found that the frequency of YAP^+ in Estonians is 7 %, close to that typically found in northern Europeans and lower than that in southern Europeans (Hammer et al. 1997). More importantly, Estonian YAP^+ chromosomes have predominantly (~73 %) type A *DYS19* repeat (186 bp), corresponding to a combined *YAP/DYS19* haplotype, not found in a divergent and large sample of East and South-East Asians, but present in Europeans and North Africans (Hammer et al. 1997). It suggests that (most of) Estonian YAP^+ paternal lineages are possibly of different (immediate) origin than those found in Asia.

Allowing a little joking, it means that besides "Siberian fathers", Estonians (and very likely also Finns and Saami) have also a fraction of Caucasoid paternal lineages, shared with Indo-European and specifically with Hamito-Semitic language group speakers of northern Africa.

Returning to the C allele: a possible earlier origin of this point mutation would make our interpretation easier where we consider that perhaps as many as 10 million Russians have the C allele. One needs some idea to explain from where Russians received their C allele. In this context it was important to find the frequency of *DYS7C* and the C allele among linguistically very close Central European Slavic people: Slovaks and Czechs. It turned out that the frequency of *DYS7C* in Slovaks and Czechs is by far lower than among Russians: 2,8 %-3 % *(V. Ferak,* personal communication). It strongly suggests that the presence of the C allele was not characteristic for proto-Slavic people, but it was acquired by admixture by proto-Russians, migrating to eastern parts of Europe some 1300 YBP. And acquired from people living in Eastern Europe at that time - most likely from proto-Finno-Ugric populations. This scenario seems more likely than any later admixture with Mongols during their invasion, which started in 13th century and much more likely than even very late admixture with Buryats and Yakut, becoming possible after Russian Empire conquered Siberia starting from 16th century. In fact, the last speculation is totally out of question. Although the details of the conquest of Siberia lend some credence to possible significant male gene flow from Russians to Yakut and Buryats, it is not very likely that this flow against the gradient actually worked here: the C allele frequency among contemporary Yakut is 4-5 times of that among Russians.

Summing up this chapter, one can conclude that the finding that there is a point mutation, shared at a high frequency by a few small Siberian-Mongolian populations and European Finno-Ugric speaking people both at the Baltic and Volga area, is most intriguing. The other significant finding is that this mutation seems to be virtually nonexistant in western Europe and also among Slovaks and Czechs, but found at relatively high frequency among Russians, otherwise genetically and linguistically close to the former. It is likely therefore that proto-Russians received this Y-chromosomal marker by admixture from living in eastern Europe proto-Finno-Ugric populations.

Analysis of classic genetic polymorphisms suggests that the split between proto-Mongoloids and proto-Caucasoids occurred about 55 000 YBP (*Nei* & *Royshoudhury*, 1993). Using mtDNA polymorphisms, the initial radiation of Mongoloids was dated as a much earlier event (*Ballinger* et al. 1992). These estimates, however, are not of any direct use here. Siberian mtDNA lineages divergence estimates vary from –13 000–52 000 YBP and other data reveal that Mongolia is inhabited by modem humans at least for some 20 000–25 000 years *(Fiedel,* 1992). Besides very large ambiguities of the data given above, there are indications of specific male gene long-distance movements in Paleolithic (*Hammer* et al. 1997, *Deka* et al. 1996; but see also *Cavalli-Sforza*, 1997), perhaps not coinciding with that of females. What is absent at present is a systematic study of the occurrence of the C allele among these numerous populations who occupy the geographic space between the Volga basin and Lake Baikal–East Siberia.

Conclusions

Irrespective whether we analyse maternal or paternal lineages of the Finno-Ugric people, we end up with a finding that the northeastern European Finno-Ugric speaking people, including Saami, are similar indeed. While in case of maternal lineages this similarity is shared, in surprisingly detailed way, by other European Caucasoids, the aforementioned distinct Y-chromosomal mutation, present at a high frequency in northeastern Finno-Ugric languages speaking males, seems to be totally absent in the western part of Europe. Its moderate frequency among eastern Slavic populations and high incidence in some Siberian populations are the facts deserving detailed attention from the point of view of demographic history of populations, speaking Finno-Ugric, Uralic in general, as well as Altaic languages.

It is clear that the Finno-Ugrians share their maternal lineages with other Caucasoids and not with Mongoloids, at least in any larger extent. Can we find, inside this Pan-European homogeneity of mtDNA haplogroups, certain Finno-Ugric variants? We think that it is possible. Not necessarily Finno-Ugric, but certainly regional. As an attempt, we analysed recently –2100 Caucasians, including a large Estonian sample, for the 9-bp deletion in the intergenic COII/ $tRNA^{Lys}$ region. We ended up with 20 deletion mutants and seven triplications of this 9-bp motif. One of the varieties of the deletion mutants was a family of haplotypes at the background of the European-specific RFLP haplogroup T with a characteristic two-basepair upstream shift in the place of the excision of the 9-bp motif. Since this specific variety of deletion mutation found in Estonians is shared by some of our neighbours, not necessarily Finno-Ugric speakers only, it can serve as a regional marker. The other deletion variety at the background of haplogroup H was further characterised by two point mutations in the flanking sequences. What we suggest here is that this kind of additional mutations and specifically their varieties are unique enough to trace detailes of maternal inheritance at the inter-European level.

Since the similarities both in maternal and paternal lineages are shared between Finns, Estonians and the Volga basin Finno-Ugric speakers on one hand, and Saami on the other, we

wish to question traditional stressing of the genetic "outliemess" of the Saami population. So far as we speak about the Y chromosome and mtDNA of the Saami population, the differences are more easily explained by drift than by any extensive Mongoloid influence. Our results allow also to question the origin of the DYS7C-linked C allele of the Y chromosome and to suggest that it has first occurred in Finno-Ugric population and only considerably later found its way to (some) Siberian populations.

Lineage analysis, at present specifically based on the mitochondria1 and the Y chromosomal DNA polymorphisms, is extraordinarily powerful tool for detailed analysis of the genetic history of humans. The era has just started and only a fraction of the heuristic power of the approach is used so far. But even this fraction can be considerably more illuminating provided true interdisciplinary spirit of the enquiry is achieved.

There is always a great temptation to exceed the accumulation of experimental data and to bring up new wild hypotheses. We think that one should be tolerant here: these hypotheses, irrespective of their final validity, do serve as hints to where the "fieldwork" should concentrate at a given time. And to concentrate we unfortunately must in this world of very limited resources for scientific exploration.

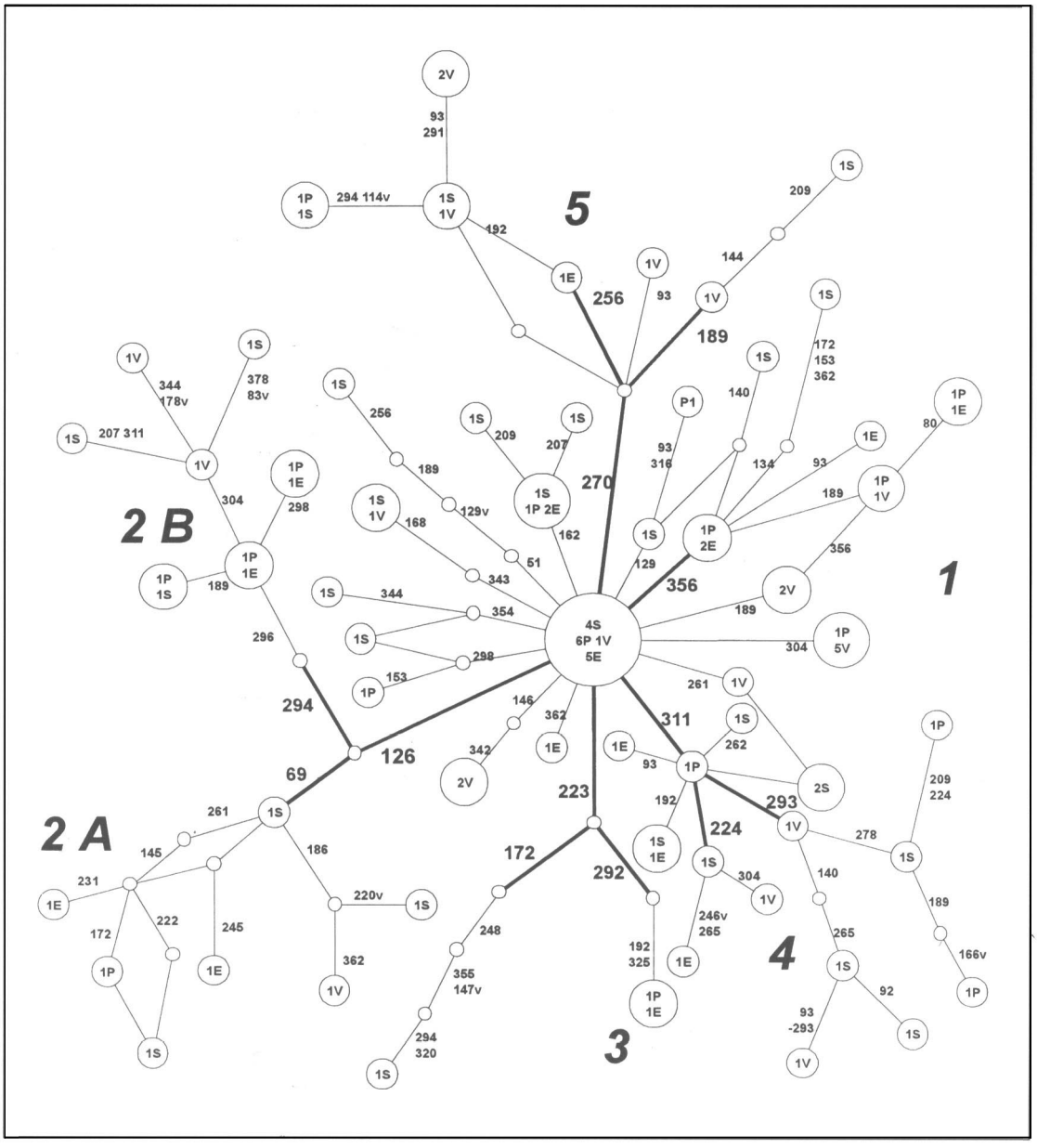

***Figure** 1. Median network of* 100 *Estonian mitochondrial DNA lineages, based on their D-loop HVR* I *sequence. Haplogroup nomenclature according to Richards et al.* (1996). *Central nod corresponds to the Cambridge Reference Sequence (CRS; Anderson et al.* 1981). *Geographic localization of individuals: V – Viljandi, Central Estonia; P – Põlva, South-Eastern Estonia; S – Saaremaa, Western Archipelago; E – other areas; All three-digit numbers correspond to mutated positions relative to the CRS; for simplicity, only three last digits are used, so that e.g.* 126 *corresponds to nucleotide* 16126 *in mtDNA, etc.*

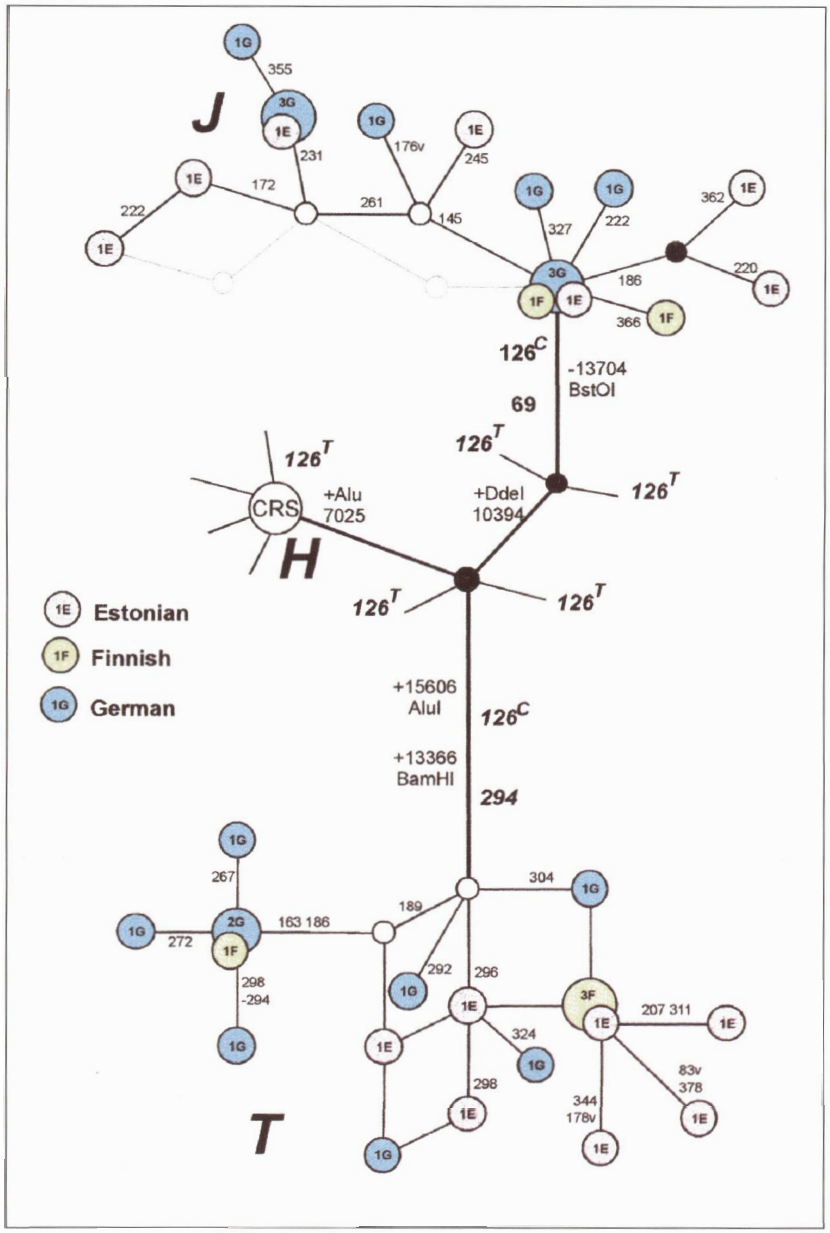

Figure 2. *An illustration of a reduced combined median network, where both the HVR I sequence data (as in Fig. 1) and RFLP data are used. Haplogroup nomenclature according to Torroni et al. (1996). Three-digit numbers as in Fig. 1, four- and five-digit numbers correspond to actual nucleotide numbers in CRS. Plus and minus define the gain and the loss of the corresponding restriction enzyme cleavage sites (**AluI**, **BamHI** etc.). E – Estonians (our data); F – Finns (Sajantila et al. 1995); **G** – Germans (Richards et al. 1996, taking haplogroup J identical to haplogroup 2A and haplogroup T identical to 2B).*

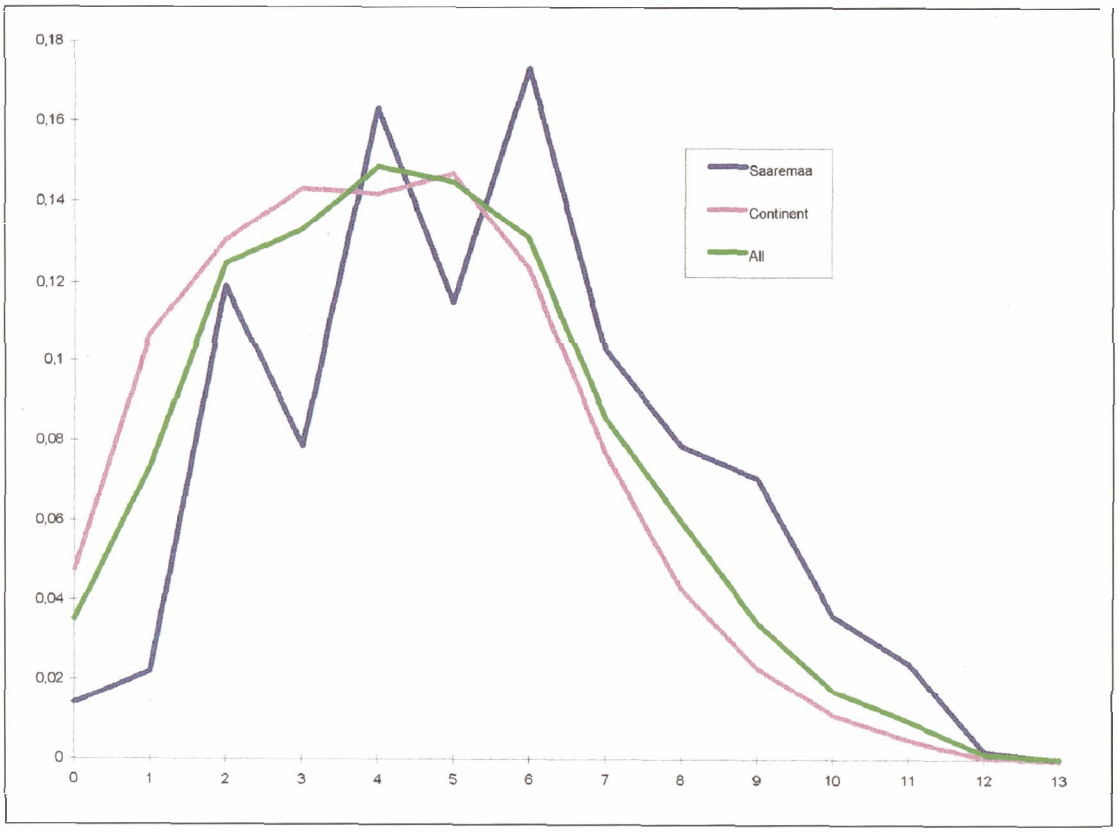

Figure 3. *Pairwise mismatch distribution of Estonian mitochondria1 DNA lineages. For further explanations see text and Table 1.*

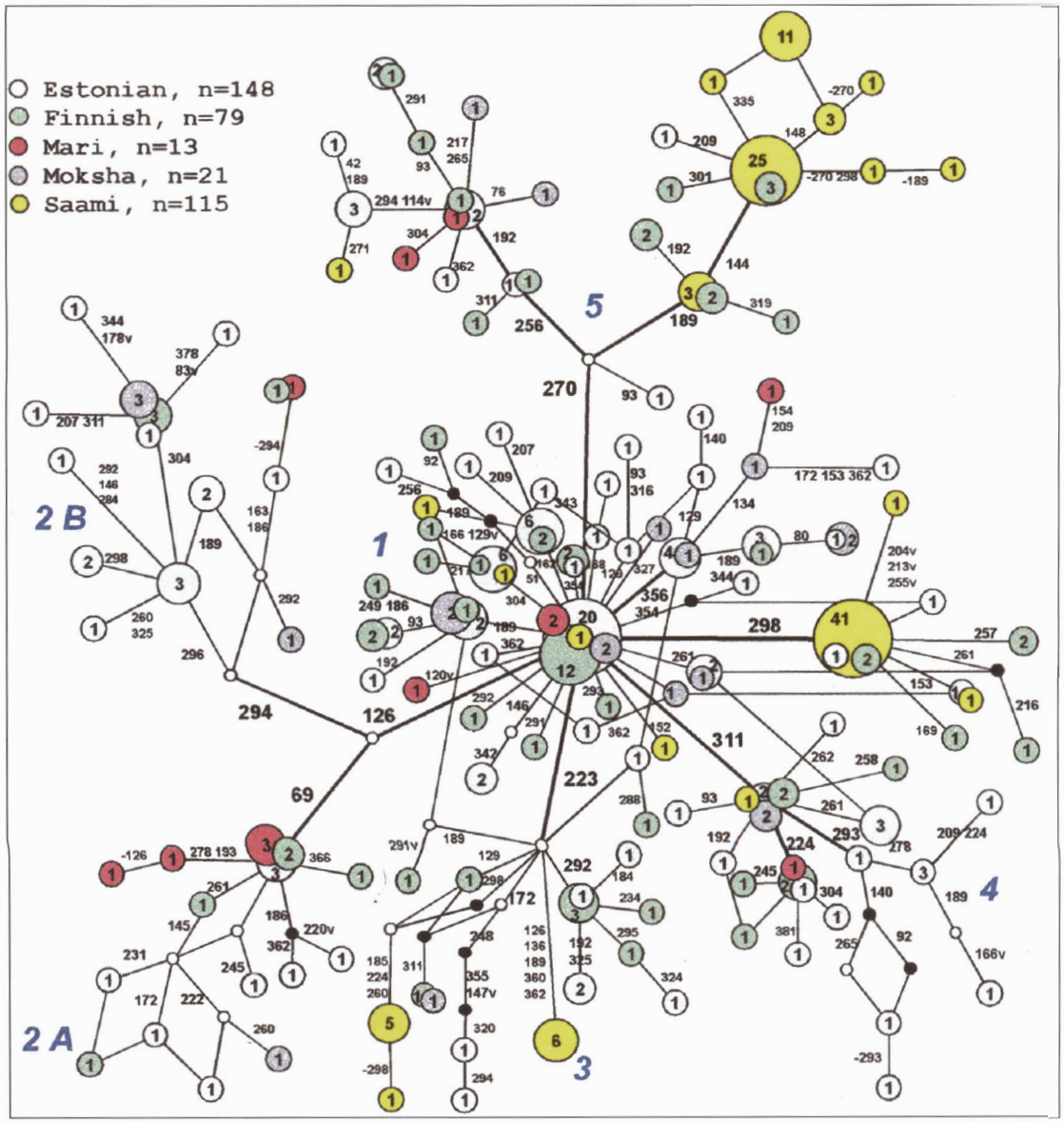

***Figure** 4. Mitochondrial D-loop HVR I – based median network of some North-European Finno-Ugric speaking populations. Abbreviations for variant positions as in Fig. 1. Numbers in the bars correspond to the number of individuals found with the corresponding DNA sequence. Sequence data from our work and from Sajantila et al. (1995). Please note the clustering of Saami compared to a more uniform distribution of other populations.*

References

Anderson, S. – Bankier, A. T. – Barrell, B. G. – De Bruijn, H.M.L. – Coulson, A.R. – Drouin, J. – Eperson, I. C., et al., 1981: *Sequence and organization of the human mitochondrial genome.* Nature 290, pp. 457465.

Ankel - Simons, F. – Cummins, J.M., 1996: *Misconceptions about.mitochondria and mammalian fertilization: Implications for theories on human evolution.* Proc. Natl. Acad. Sci. USA 93, pp. 13859–13863.

Ballinger, S. W. – Schurr, T. G. – Torroni, A. – Gan, Y. -Y. – Hodge, J. A., et al., 1992: *Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations.* Genetics 130, pp. 139–152.

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

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

Cavalli-Sforza, L. L., 1997: *Genes, peoples, and languages.* Proc. Natl. Acad. Sci. USA 94, pp. 7719–7724.

Cavalli-Sforza, L.L. – Menozzi, P. – Piazza, A., 1994: *The History and Geography of Human Genes.* Princeton Univ. Press, Princeton.

de la Chapelle, A., 1993: *Disease gene mapping in isolated human populations: the example of Finland.* J. Med. Genet. 30, pp. 857–865.

Deka, R. - Jin, L. - Shriver, M.D. - Yu, L.M. - Saha, N. – Barrantes, R. – Chakraborty, R. – Ferrell, R. E., 1996: *Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations.* Genome Res. 6, pp. 1177–1184

Diamond, J., 1997: *Guns, Germs and Steel.* Norton, New York.

Dorit, R. L. – Akashi, H. – Gilbert, W., 1995: *Absence of polymorphism at the ZFY locus of the human Y chromosome.* Science 268, pp. 1183–1185

Fiedel, S. J., 1992: *Prehistory of the Americas.* Cambridge University Press, Cambridge.

Hammer, M. F., 1994: *A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies.* Mol. Biol. Evol. 11, pp. 749–761

Hammer, M. F., 1995: *A recent common ancestry for human Y chromosomes.* Nature 378, pp. 376–378.

Hammer, M.F. – Spurdle, A. B. – Karafet, T. – Bonner, M. B. – Wood, E. T. – Novelletto, A. – Malaspina, P. – Mitchell, R. J. – Horai, S. - Jenkins, T. - Zegura, L.S., 1997: *The geographic distribution of human Y chromosome variation.* Genetics 145, pp. 787–805.

Harpending, H.C. – Sherry, S.T. – Rogers, A. R. – Stoneking, M., 1993: *The genetic structure of ancient human populations.* Curr. Anthropol. 34, pp. 483496

Hirszfeld, L. – Hirszfeld, H., 1919: *Essai d'upplication des mhthodes au probleme des races.* Anthropologie 29, pp. 505–537.

Jobling, M.A. – Samara, V. – Pandya, A. – Fretwell, N. – Bernasconi, B. – Mitchell, R. J. – Gerelsaikhan, T. – Dashnyam, B. – Sajantila, A. – Salo, P. J. – Nakahori, Y. – Disteche, M.C. – Thangaraj, K. – Singh, L. – Crawford, M. H. – Tyler-Smith, C., 1996: *Recur-rent duplication and deletion polymorphism on the long arm of the Y chromosome in normal males.* Hum. Molec. Genet. 5, pp. 1767–1775.

Jobling, M. A. – Tyler-Smith, C., *1995:Father.s and sons: the Y chromosome and human evolution.* Trends Genet. 11, pp. 449456.

Lahermo, P. – Sajantila, A. – Sistonen, P. – Lukka, M. – Aula, P. – Peltonen, L. – Savontaus, M. -L., 1995: *The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA.* Am. J. Hum. Genet. 58, pp. 1309–1322.

Mountain, J. L. – Cavalli-Sforza, L. L., 1994: *Inference of human evolution through cladistic analysis of nuclear DNA restriction polymorphisms.* Proc. Natl. Acad. Sci. USA 91, pp. 6515–6519.

Nei, M. – Roychoudhury, A. K., 1993: *Evolutionary relationship of human populations on a global scale.* Mol. Biol. Evol. 10, pp. 927–943.

Passarino, G. - Semino, O. - Bernini, L.F. - Santachiara - Benerecetti, A. S., 1996a: *Pre-Caucasoid and Caucasoid genetic features of Indian population revealed by mtDNA polymorphism* ~Am. J. Hum. Genet. 59, pp. 927–934.

Passarino, G. – Semino, O. – Modiano, G. – Bernini, L.F. – Santachiara - Benerechetti, A. S., 1996b: *mtDNA provides the first known marker distinguishing proto-Indians from the other Caucasoids; it probably predates the diversification between Indians and Orientals.* Ann. Hum. Biol. 23, pp. 121–126.

Rogers, A. R. – Harpending, H., *1992: Population growth make waves in the distribution of pairwise genetic distances.* Mol. Biol. Evol. *9,* pp. 552–*569.*

Richards, M. – Corte-Real, H. – Forster, P. – Macaulay, V. – Wilkinson-Herbots, H. – Demaine, A. – Papiha, S. – Hedges, R. – Bandelt, H.-J. – Sykes, B., *1996: Paleolithic and Neolithic lineages in the European mitochondrial gene pool. Am-* J. Hum. Genet. *59,* pp. 185–203.

Relethford, J. H. – Harpending, H. C., *1995: Ancient differences in population size can mimic a recent African origin of modern humans.* Current Anthrop. *36,* pp. 667–674.

Sajantila, A. – Paabo, S., *1995: Language replacement in Scandinavia.* Nature Genetics *11,* pp. *359–360.*

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

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

Spurdle, A. B.-Jenkins, T., 1992: *The Y chromosome as a tool for studying human evolution.* Curr. Opinion Genet. Developm. 2, pp. 487–491.

Tavaré, S. – Balding, D. J. – Griffith, R. C. – Donnelly, P., *1997: Inferring coalescence times from DNA sequence data.* Genetics *145,* pp. 505–518.

Torroni, A. – Lott, M. T. – Cabell, M. F. – Chen, Y.-S. – Lavergne, L. – Wallace, D. C., *1994: mtDNA and the origin of Caucasoids: 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,* pp. *760–776.*

Torroni, A. – Huoponen, K. – Francalacci, P. – Petrozzi, M. – Morelli, L. – Scozzari, R. – Obinu, D. – Savontaus, M.-L. – Wallace, D. C., *1996: Classification of European mtDNAs from an analysis of three European populations.* Genetics *144,* pp. *1835–1850.*

Vigilant, *L. M.* – Stoneking, H. – Harpending, H.-Hawkes, K.-Wilson, A. C., *1991: African population and the evolution of human mitochondria/ DNA.* Science *253,* pp. 1503–1507.

Wakeley, J. – Hey, J., *1997: Estimating Ancestral Population Parameters.* Genetics *145,* pp. 847–855.

Whitfield, L. S. – Sulston, J. E. – Goodfellow, P. N., *1995: Sequence variation of the human Y chromosome.* Nature *378,* pp. *379–380.*

Zerjal, T. – Dashnyam, B. – Pandya, A. – Kayser, M. – Roewer, L. – Santos, F. R. – Schienhovel, 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. – Tyler-Smith, C., *1997: Genetic relationships of Asians and Northern Europeans, revealed by Y-Chromosomal DNA analysis.* Am. J. Hum. Genet. *60,* pp. *1174–1 183.*