

Migration Waves to the Baltic Sea Region

T. Lappalainen¹, V. Laitinen², E. Salmela^{1,3}, P. Andersen⁴, K. Huoponen², M.-L. Savontaus² and P. Lahermo^{1*}

¹*Finnish Genome Center, Institute for Molecular Medicine Finland, University of Helsinki, Haartmaninkatu 8, P.O. Box 63,00014 University of Helsinki, Finland*

²*Department of Medical Genetics, University of Turku, Känkämyllykatu 10, 20520 Turku, Finland*

³*Department of Medical Genetics, University of Helsinki, Haartmaninkatu 8, P.O. Box 63, 00014 University of Helsinki, Finland*

⁴*Department of Neurology, Umeå University Hospital, University of Umeå, 901 85 Umeå, Sweden*

Summary

In this study, the population history of the Baltic Sea region, known to be affected by a variety of migrations and genetic barriers, was analyzed using both mitochondrial DNA and Y-chromosomal data. Over 1200 samples from Finland, Sweden, Karelia, Estonia, Setoland, Latvia and Lithuania were genotyped for 18 Y-chromosomal biallelic polymorphisms and 9 STRs, in addition to analyzing 17 coding region polymorphisms and the HVS1 region from the mtDNA. It was shown that the populations surrounding the Baltic Sea are genetically similar, which suggests that it has been an important route not only for cultural transmission but also for population migration. However, many of the migrations affecting the area from Central Europe, the Volga-Ural region and from Slavic populations have had a quantitatively different impact on the populations, and, furthermore, the effects of genetic drift have increased the differences between populations especially in the north. The possible explanations for the high frequencies of several haplogroups with an origin in the Iberian refugia (H1, U5b, I1a) are also discussed.

Keywords: Y chromosome, mitochondrial DNA, Baltic, Northern Europe, population history

Introduction

Analysing and separating a variety of migratory and demographic events from the current genetic variation of a region is a challenging task. In human population genetic studies that aim to dissect population history, analysis has traditionally consisted of pooling genetic data for each population and comparing the populations using various statistical methods. However, it has been shown that this approach is not entirely unproblematic since differences between haplogroups usually explain a larger proportion of the total variance than differences between populations (Bosch *et al.* 1999). Thus, having solely populations as the main unit of study without a combination of lineage-based information may lead to a loss of information and a failure to identify more subtle trends. Indeed, phylogeographic analyses of haplogroup variation have greatly increased our understanding of population history on

a wide geographical and temporal scale, and interpretation of haplogroup frequencies among populations relies heavily on the knowledge of the historical strata that different haplogroups represent.

Throughout their history, the populations of the Baltic Sea region have been affected by migrations from both Western and Central Europe and from the east. The region was first settled both from the south-east and from the south soon after the retreat of the continental ice sheet some 12000 years ago. The first ceramic culture in Scandinavia was of southern origin, whereas the eastern and northern shores of the Baltic Sea were affected by the Comb Ceramic culture that may have a Finno-Ugric association. These early Neolithic cultures were followed by the Corded Ware and Bronze Age cultures that affected Northern Germany, Scandinavia, the Baltic countries and coastal Finland. The Northeastern region, however, had close ties to the Russian area especially in the Bronze and Early Iron Age. (Huurre 1990, Siiriäinen 2003) Later political and population histories of the countries are intertwined with Swedish and Russian influence over large areas. Furthermore, the Germans have had a prominent role in urban life especially in the Baltic states (Alenius 2000). At present, the linguistic variation in the Baltic Sea region is substantial, with four

*Corresponding author: Päivi Lahermo, Finnish Genome Center, Institute for Molecular Medicine Finland, University of Helsinki Haartmaninkatu 8, P.O. Box 63,00014 University of Helsinki. Tel. +358-9-191 25476, Fax. +358-9-191 25478, E-mail: paivi.lahermo@helsinki.fi

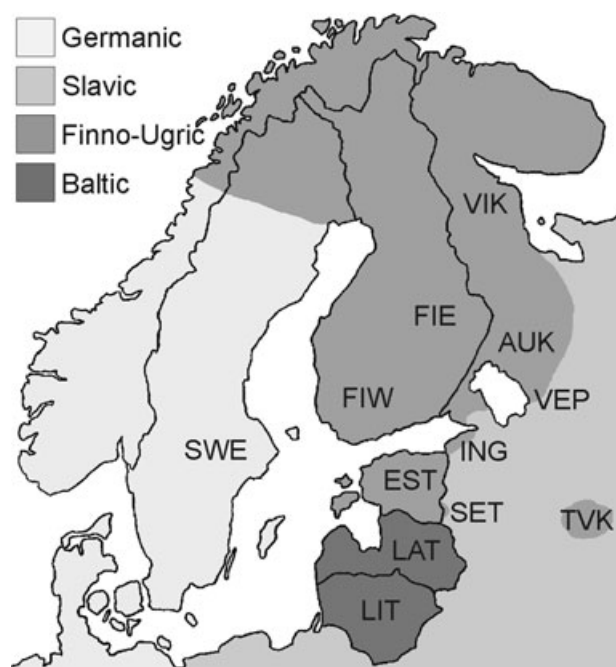


Figure 1 Map of the studied populations and their linguistic groups. For the abbreviations, see Figure 2.

major language groups: Finno-Ugric, and the Indo-European branches Baltic, Germanic, and Slavic (Fig. 1, Table 1).

In this study, we analyzed the population history of the Baltic Sea region. To achieve this, we studied the variation

of both mitochondrial DNA and the Y chromosome with a special emphasis upon characterising the regional phylogeography of the most common haplogroups. The analysis covered both the internal variation of the region as well as the immigrations to the Baltic Sea area.

Material and Methods

Samples and Genotyping

Blood samples were collected from healthy unrelated individuals from populations of the Baltic Sea region (Fig. 1, Table 1). All samples were collected according to the principles of the Declaration of Helsinki (1964), and the project was approved by the local ethical committees. The Swedish samples are mainly from eastern Sweden. To acquire adequate sample sizes, the Aunus, Tver and Viena Karelians, Vepsians, Ingrians and 44 additional Karelian samples were pooled together as Karelians for the Y-chromosomal analysis and some mtDNA calculations. Respectively, the Estonians, Latvians, Lithuanians and Seto were pooled as Balts in some analyses. The Seto are not included in the Y-chromosomal analysis due to their small number of male samples. Even though these arrangements may cause a loss of resolution or even a small bias, pooling closely related populations is still preferable to exclusion of a mass of data. A subset of the Finnish and Baltic Y-chromosomal data has been presented before, (Lappalainen *et al.* 2006, Laitinen *et al.* 2002) and the Finnish (Finnilä *et al.*

Table 1 The Y-chromosomal haplogroup frequencies (%) within the populations, and their linguistic affiliations

Language	Estonian	Latvian	Lithuanian	Karelian	Eastern Finnish	Western Finnish	Swedish
	Finno-Ugric	IE ¹ /Baltic	IE ¹ /Baltic	Finno-Ugric	Finno-Ugric	Finno-Ugric	IE ¹ /Germanic
Y* (xC,DE,F)	0	0	0	0	0	0.4	0
C	0	0	0	0	0	0.4	0
DE	2.5	0.9	1.2	2.3	0	0.9	1.3
F* (xI,J,K)	0.8	0.9	1.8	0.8	0.3	0.4	4.4
I1a	11.9	3.5	4.9	15.2	19	40	35.6
I1b	4.2	2.7	4.9	2.3	0	0.4	0
I1c	0.8	0.9	1.8	0	0.7	0.9	1.9
J	1.7	0	1.8	0	0	0	0
K* (xN,P)	0.8	0	0	0	0	0	0.6
N* (xN2,3)	0	0	0	0	0	0.4	0
N2	0	0	0	0.8	0	0.9	0
N3	33.9	41.6	43.9	53	70.9	41.3	14.4
P* (xQ,R)	0.8	0	0.6	0	0	0	4.4
Q	0.8	0.9	0	0	0.3	0	0
R* (xR1a,R1b)	0	0	0	0	0.3	0	0
R1b	4.2	9.7	4.9	0.8	2.6	5.2	13.1
R1a1	37.3	38.9	34.1	25	5.9	8.7	24.4
n	118	113	164	132	306	230	160

¹IE: Indo-European

2001) and Russian (Malyarchuk & Derenko 2001, Loogväli *et al.* 2004) mitochondrial DNA data were collected from the literature. Because the total mtDNA sequencing by Finnilä *et al.* (2001) has been selective in favour of some haplogroups, a part of the sequenced samples were randomly excluded until the mtDNA data matched the true haplogroup frequencies given in their Table 1.

A total of 1223 male samples were genotyped for 18 Y-chromosomal biallelic polymorphisms (SRY-1532, M216, M203, P14, M170, M253, P37, M223, 12f2, M9, LLY22g, P43, Tat, M45, P36, M207, P25, M17) mostly on the Sequenom[®] platform, and for nine STR loci (DYS19, DYS385a/b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393). From the mtDNA, a total of 1269 samples were analyzed for 17 SNPs in the coding region (663, 3010, 4529, 4580, 4769, 4833, 6776, 7028, 7309, 8994, 9055, 12308, 13263, 13368, 13708, 14470, 5178). The Karelian and Seto samples and SNP A13263G from the whole sample set were genotyped by RFLP analysis, and the Swedes and Balts were analyzed on the Sequenom[®] platform. In addition, hypervariable segment 1 (HVS1) was sequenced in all the samples with standard methods. Details of the genotyping and haplogroup assignment are given in Supplementary Table 1.

Statistical Analysis

Y-chromosomal haplogroups were constructed from the biallelic marker data according to common guidelines (Y Chromosome Consortium 2002). MtDNA haplogroups were assigned based on both coding region and HVS1 polymorphisms (Finnilä *et al.* 2001, Loogväli *et al.* 2004, Richards *et al.* 1998, Macaulay *et al.* 1999, Kivisild *et al.* 2002, Achilli *et al.* 2004,2005). The frequencies of haplogroups were calculated in each of the studied populations.

We performed phylogeographic analysis for Y-chromosomal haplogroups N3, I1a and R1a1, and mtDNA haplogroups H and U, since these haplogroups had a sufficient number of samples for the analysis of patterns of intrahaplogroup variation. In these mtDNA analyses all the Finnish samples sequenced belonging to the haplogroups H and U were used (Finnilä *et al.* 2001). The Network 4.112 software was used to construct median-joining networks of the intrahaplogroup variation (Bandelt *et al.* 1995, 1999). The Y-chromosomal STR markers were assigned weights in inverse accordance with their variation within the haplogroup. The networks were simplified by partially removing circular structures according to the weighting scheme, and excluding the singleton haplotypes in the case of the very complex N3 network – however, the singletons are naturally included in the calculation of the coalescence age. In the mitochondrial DNA analysis, the coding region

polymorphisms were given a very high weight, most HVS1 polymorphisms were allocated an intermediate one, and those HVS1 sites with a mutation rate over 3 times the average of HVS1 were downweighted (Meyer *et al.* 1999). The mtDNA position 16519 was omitted from network and coalescence age calculations due to its hypervariable status. The relative sample sizes of different populations were calculated to ease the interpretation of the network figures drawn from population samples of varying size. Coalescence ages for the haplogroups were also calculated by the Network 4.112 software. The mtDNA coalescence analysis was based on HVS1 (16090–16365) variation with a mutation rate of 1 transition per 20180 years (Forster *et al.* 1996). For Y-chromosomal coalescence estimates a rate of 1 mutation per 3623 years was used (Zhitovtovsky *et al.* 2004). Also Bayesian-based calculations for coalescence age were performed (Wilson *et al.* 2003), but they showed such a lack of robustness for small deviations of the data that the network approach was preferred for the final analysis.

Additional analyses were performed for the Y-chromosomal and mtDNA data of each population. We used the Arlequin 2.001 software (Schneider *et al.* 2000) for calculations of haplotype diversity (Nei 1987), mean number of pairwise differences, and genetic distances measured as F_{ST} and R_{ST} (Slatkin 1995). Multidimensional scaling (MDS) plots for genetic distances were constructed with R software (R Development Core Team 2005). SAMOVA 1.0 software (Dupanloup *et al.* 2002) was used to assess genetic grouping of populations. Admixture proportions were calculated by Admix 2.0 (Dupanloup & Bertorelle 2001) from the Y-chromosomal data of selected populations by calculating molecular distance as squared difference in allele size and bootstrapping with 50000 replicates. The diversity and SAMOVA analyses were performed also for each major haplogroup. Furthermore, matches for the most common Y-chromosomal haplotypes in our dataset were found in the Y-chromosomal haplotype reference database (Willuweit & Roewer 2007, www.yhrd.org).

Results

Population-based Y-chromosomal Analyses

The Y-chromosomal data is given in Supplementary Table 2, and the haplogroup frequencies in Table 1. In the MDS plot of the genetic distances calculated from the total Y-chromosomal data (Fig. 2a), there were clusters consisting of the Balts, Karelians/Eastern Finns, and Swedes/Western Finns, which accounts for 8.0% of the variation. The populations in the Baltic states had a very high haplotype diversity up to 0.998, while the diversity was lower in the northeastern populations (Table 2). In the admixture analysis of the Estonians, the parental populations of Latvians+Lithuanians and

Table 2. The haplotype diversities and coalescence ages of the total Y-chromosomal data and the major haplogroups

	n	Total Haplotype diversity	n	I1a Haplotype diversity	n	N3 Haplotype diversity	n	R1a1 Haplotype diversity
Estonian	118	0.998 ± 0.002	14	0.956 ± 0.045	40	0.994 ± 0.008	44	0.992 ± 0.007
Latvian	113	0.992 ± 0.003	4	NA	47	0.962 ± 0.013	44	0.994 ± 0.006
Lithuanian	164	0.990 ± 0.004	8	1.000 ± 0.063	72	0.936 ± 0.020	56	0.990 ± 0.006
Karelian	132	0.966 ± 0.009	20	0.895 ± 0.043	70	0.899 ± 0.030	33	0.936 ± 0.027
Eastern Finnish	306	0.966 ± 0.006	58	0.968 ± 0.011	217	0.934 ± 0.013	18	0.994 ± 0.021
Western Finnish	230	0.974 ± 0.006	91	0.880 ± 0.030	95	0.960 ± 0.009	20	0.990 ± 0.019
Swedish	160	0.989 ± 0.003	57	0.939 ± 0.021	23	0.917 ± 0.040	39	0.991 ± 0.008
TMRCA (ky)	1223	—	252	7.7 ± 1.3	564	8.8 ± 1.5	254	10.7 ± 1.4

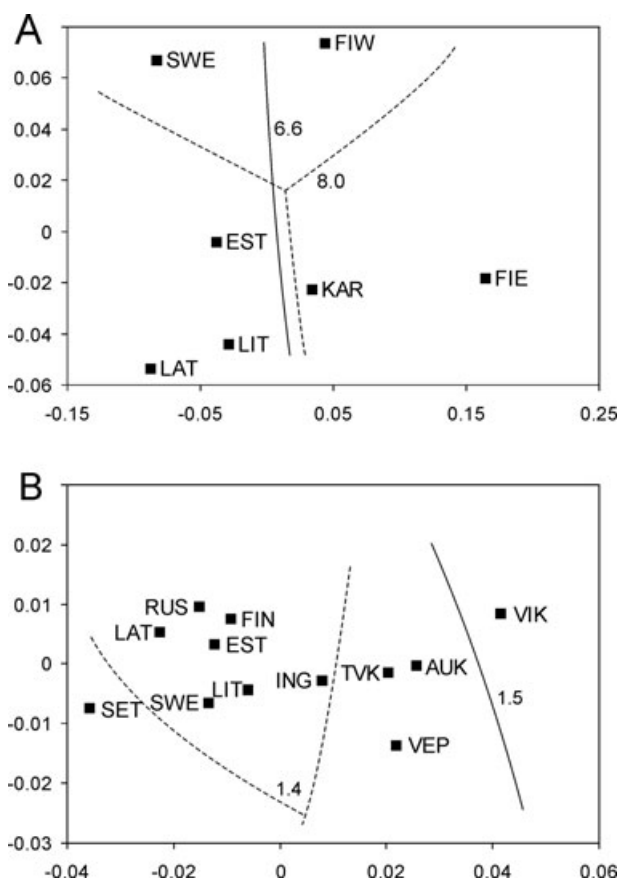


Figure 2 MDS plot of genetic distances calculated from the total Y-chromosomal data (stress 0.27) (A), and of total mtDNA data (stress 0.19) (B). The solid lines denote a division into two groups and the dotted line a division into three groups in SAMOVA analysis, and the numbers denote the respective percentages of variance. Abbreviations: Estonian (EST), Latvian (LAT), Lithuanian (LIT), Seto (SET), Karelian (KAR), Aunus Karelian (AUK), Viena Karelian (VIK), Tver Karelian (TVK), Vepsian (VEP), Ingrian (ING), Finnish (FIN), Eastern Finnish (FIE), Western Finnish (FIW), Swedish (SWE), and Russian (RUS).

Eastern+Western Finns had respective admixture coefficients of 0.81 ± 0.20 and 0.19 ± 0.20 . For the other populations the definition of parental populations was too complex and the results too weak to yield good estimates of admixture coefficients.

Y-chromosomal Haplogroups

Haplogroup N3 was much more common on the eastern side of the Baltic Sea than in Sweden (χ^2 $p < 0.001$) (Table 1). In the STR network (Fig. 3a) many haplotypes are highly specific for either Finno-Ugric or Baltic-speaking populations, which is supported by the SAMOVA analysis, in which the grouping into Latvians and Lithuanians vs. the others explains as much as 20.3% of the variation. The Estonians harbor both Baltic and Finnic haplotypes, and they had the highest haplotype diversity (Table 2). The age of the haplogroup was 8,800 years.

Haplogroup I1a had a high frequency of up to 40% in Sweden and Western Finland and intermediate frequencies in the other Finno-Ugric populations, while it was almost absent among Latvians and Lithuanians (Table 1). In the network the Finnish and Swedish haplotypes appear to be separate (Fig. 3b), which is supported by the SAMOVA analysis that separated the Swedes/Balts from the others with a moderate 6.9% of variation among these groups. The highest diversities were in Eastern Finland, the Baltic states, and Sweden (Table 2). The age of the haplogroup was 7,700 years (Table 2).

Haplogroup R1a1 had high frequencies up to 39% among all the populations with the exception of the Finns (Table 1). In the very starlike network (Fig. 3c), the Karelians exhibited a limited diversity of haplotypes, which is consistent with the haplotype diversity calculations (Table 2) and SAMOVA analysis, where the grouping of the Karelians *versus* the others explains 9.0% of the variation. The age of the haplogroup was as high as 10,700 years (Table 2).

The frequencies of the other haplogroups were low and mostly lacked strong geographical patterns. R1b was common

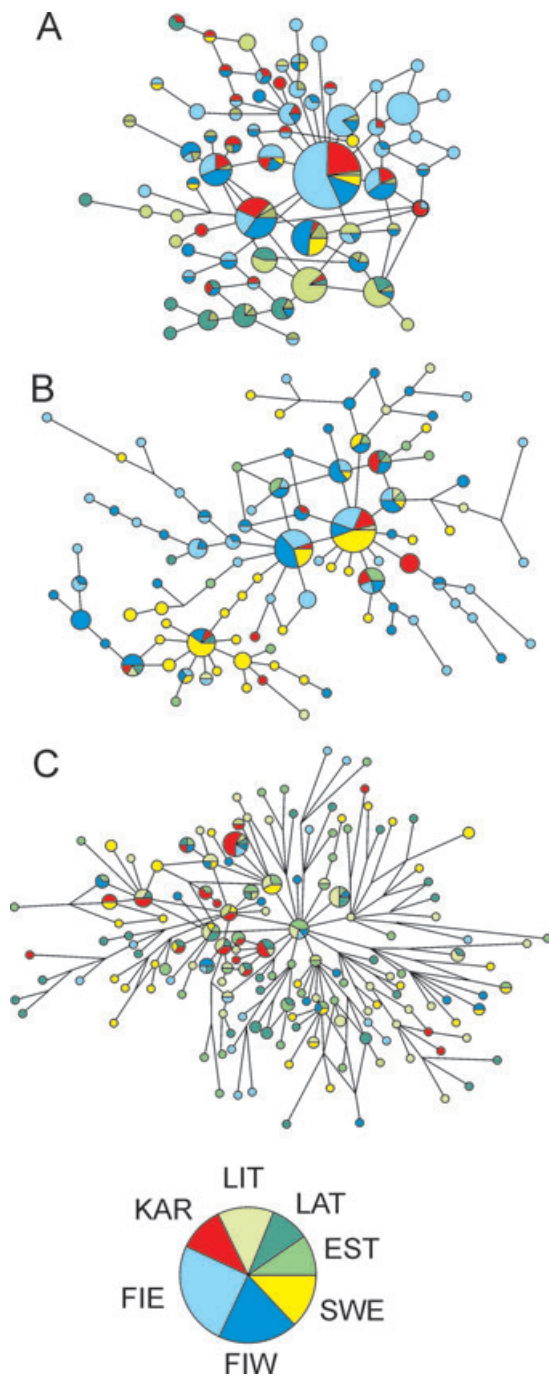


Figure 3 Median-joining networks of the haplotype variation of Y-chromosomal haplogroups N3 (A), I1a (B), and R1a1 (C). The lengths of the branches correspond to the number of STR mutations, and the sizes of the nodes denote the number of samples of the corresponding haplotype. Note that the singleton haplotypes have been excluded from the N3 network. The pie chart denotes the proportions of the different populations in the total sample set of the Y-chromosome. For the abbreviations, see Figure 2.

in Sweden and in Western Finland, whereas I1b was more abundant in the Baltic states and Karelia.

Population-based Mitochondrial DNA Analyses

The full mitochondrial DNA data is given in Supplementary Table 3, and the mtDNA haplogroup frequencies are given in Table 3. The MDS plot of the genetic distances calculated from the total mtDNA data had only one cluster, but with exclusively the Karelian populations on its other side, a pattern mainly supported by the SAMOVA analysis (Fig. 2b). The mean pairwise differences were relatively similar across the populations with slightly lower numbers among the Karelians and Russians (Table 4).

Mitochondrial DNA Haplogroups

Haplogroup H was very common among all the populations, but with considerable variation in the subhaplogroup frequencies (Table 3). H1* was common among the Karelians, Swedes, and some Baltic populations with frequencies up to 18%, and rare especially in Finland and Estonia, while H1f was very specific to the Finns and Karelians. H3 was relatively rare, with frequencies of a few percent. The mean pairwise differences were again lower among the Karelians and Russians and, surprisingly, the highest among the Seto (Table 4). The low diversity of haplogroup H among the Karelians can be observed also in the haplotype network (Fig. 4a): compared to e.g. the Balts, a much bigger proportion of the Karelian haplotypes are in high-frequency modal haplotypes. The median-joining network of haplogroup H consists of highly starlike clusters, and the haplotypes tend to have either very high or very low frequencies. The differences in subhaplogroup frequencies are clearly visible in the network e.g. in the case of the abundance of H2 among the Swedes. The coalescence age of haplogroup H in our dataset was 36,700 years (Table 4).

Of the U subhaplogroups, U4 was the most frequent among the Latvians, Seto, and Tver Karelians (7.1–8.8%) (Table 3). U5b and U5b1b1 were common in Karelia and especially among the Viena Karelians where U5b1b1 reached a high frequency 16.1%. The Estonians had the highest number of mean pairwise differences (Table 4). Haplogroup U had a network with some starlike clusters, too, but to a lesser degree than haplogroup H (Fig. 4b), and especially subhaplogroups U5b*, U5b1b1 and K had very common modal haplotypes, whereas the variation of U5a was scattered into haplotypes of moderate frequency. The SAMOVA analysis separated the Seto and Vepsians with small sample sizes from the main group. The coalescence age was as high as 68,400 years (Table 4).

Haplogroup Z was observed among the Finns, some Karelian populations, Russians and Swedes with a low frequency. Asian haplogroups A, C, G and D were rare in the Baltic

Table 3 The mitochondrial DNA haplogroup frequencies (%) within the populations

					Aunus	Viena	Tver					Total		
	Estonian	Latvian	Lithuanian	Seto	Karelian	Karelian	Karelian	Vepsian	Ingrian	Karelian	Finnish ¹	Swedish	Russian ²	
X	0.9	0	0	0	0	0	0	0	0	0	1.3	0.3	4.0	
Z	0	0	0	0	0	1.1	0	0	2.6	0.4	2.5	0.3	2.0	
A	0	0	0	0	0	0	0	0	0	0	0	0	0	
C	0	0	0	0	0	0	0	0	0	0	0	0.3	0	
D5	0	0	0	0	2.8	11.5	3.3	0	0	3.5	0	0	0	
D	0	0	0	0	0	0	1.6	0	0	0.2	0	0	0	
G	0	0.9	0	0	0.5	0	0	0	0	0.2	0	0	0	
other	3.4	1.8	8.6	5.4	1.4	1.1	1.6	10.9	0	2.7	0	3.3	10.0	
H5	1.7	7.9	3.1	0	3.2	0	3.3	0	0	2.3	5.1	1.3	4.0	
H6	6	0	3.1	8.9	0.5	1.1	1.6	3.1	0	1	0	3.9	0	
H	17.9	10.5	17.2	12.5	6.9	11.5	14.8	21.9	7.9	10.2	13.9	8.1	24.0	
H1a	5.1	1.8	3.7	5.4	7.3	12.6	8.2	15.6	5.3	9	0	2.3	0	
H1b	4.3	4.4	5.5	0	3.2	0	3.3	1.6	2.6	2.3	0	1.3	2.0	
H1	2.6	6.1	12.3	12.5	15.6	8	18	14.1	13.2	15	2.5	12.7	6.0	
H1f	0	0	0	0	2.3	4.6	0	0	5.3	2.1	10.1	1	0	
H2a1	2.6	1.8	1.2	3.6	4.1	0	0	0	5.3	2.3	2.5	9.1	0	
H2	0	0.9	0	0	0	0	0	0	5.3	0.4	2.5	2	2.0	
H3	0.9	1.8	0.6	0	2.8	1.1	0	1.6	5.3	2.1	2.5	3.9	0	
I1	1.7	5.3	2.5	0	0.5	0	0	1.6	0	0.4	2.5	0	0	
I	0	0	0	0	3.2	1.1	0	0	0	1.8	1.3	3.3	2.0	
J1	2.6	0	0.6	1.8	0	0	0	0	0	0	0	2.6	0	
J1a	0.9	0.9	0	0	0	0	0	0	0	0	1.3	0.3	0	
J1b1	3.4	0	1.2	1.8	0.9	0	1.6	0	0	0.6	1.3	0.3	0	
J2	0	0	0	0	0	1.1	0	0	0	0.2	0	0.3	0	
J	6.0	4.4	3.1	5.4	3.2	2.3	3.3	4.7	2.6	3.7	2.5	2.3	6.0	
T1	2.6	1.8	3.7	1.8	1.8	1.1	0	0	5.3	1.8	0	2.9	4.0	
T	11.1	7.0	6.7	3.6	0.9	1.1	1.6	1.6	10.5	2.0	2.5	7.2	14.0	
K	0.9	2.6	0.6	1.8	0.5	3.4	3.3	0	0	1.6	2.5	7.5	2.0	
U1	0	0	0	0	5	0	0	0	0	2.1	0	0	0	
U2	2.6	5.3	0	0	0.5	0	0	1.6	0	0.4	0	0.3	0	
U3	0.9	3.5	0.6	0	1.4	0	0	0	0	0.6	0	1.3	0	
U4	4.3	8.8	2.5	7.1	3.2	1.1	8.2	0	0	2.7	2.5	2.9	2.0	
U5	0	0.9	0	0	0	0	0	0	0	0.2	0	0.7	0	
U5a	2.6	4.4	5.5	12.5	2.3	1.1	0	0	5.3	1.6	3.8	4.6	4.0	
U5a1	6.8	5.3	3.7	8.9	3.2	0	4.9	1.6	0	2.1	1.3	2	4.0	
U5b	2.6	2.6	3.1	1.8	2.3	6.9	3.3	6.3	7.9	4.1	8.9	3.9	0	
U5b1b	0.9	0	0	0	8.7	16.1	8.2	9.4	2.6	9.0	5.1	2.0	0	
U	0.9	3.5	1.8	3.6	5	6.9	1.6	1.6	2.6	4.1	6.3	3.3	2.0	
V	1.7	1.8	6.1	1.8	6.4	4.6	4.9	1.6	5.3	5.5	5.1	1.3	4.0	
W	2.6	4.4	3.1	0	0.5	0	3.3	1.6	5.3	1.8	10.1	1.3	2.0	
n	117	114	163	56	218	87	61	64	38	512	79	307	50	

¹From Finnilä *et al.* (2001)²From Malyarchuk & Derenko (2001) and Loogväli *et al.* (2004)

Sea region with the exception of D5 that reached a high frequency of 11.5% in Viena Karelia (Table 3).

Discussion

The aim of this study was to unravel the paternal and maternal population history of the Baltic Sea region, both the

migrations affecting the region from outside and internal population events. To yield the maximum resolution of our extensive dataset, we chose to combine population-based analyses with regional analysis of phylogeography of the major haplogroups. Haplotype diversity, age and network calculations are conventional tools of phylogeographic analysis of lineages, but the problem of network analysis is that the method itself

Table 4 The mean pairwise differences (π) and coalescence ages of the total mtDNA data and the major haplogroups.

	Total		H		U	
	n	π	n	π	n	π
Estonian	117	7.92 +/- 3.71	48	4.03 +/- 2.05	26	7.65 +/- 3.69
Latvian	114	7.77 +/- 3.64	40	4.24 +/- 2.15	42	7.01 +/- 3.36
Lithuanian	163	7.48 +/- 3.51	76	3.98 +/- 2.01	29	5.97 +/- 2.93
Seto	56	7.52 +/- 3.56	24	4.74 +/- 2.40	20	6.28 +/- 3.11
Aunus Karelian	218	7.01 +/- 3.30	100	3.69 +/- 1.88	70	5.68 +/- 2.76
Viena Karelian	87	7.14 +/- 3.38	34	2.45 +/- 1.36	31	4.55 +/- 2.30
Tver Karelian	61	6.37 +/- 3.06	30	2.70 +/- 1.48	18	5.42 +/- 2.74
Vepsian	64	5.61 +/- 2.73	37	2.88 +/- 1.55	13	3.82 +/- 2.05
Ingrian	38	6.70 +/- 3.23	19	3.64 +/- 1.93	7	NA
Total Karelian	512	6.72 +/- 3.18	240	3.27 +/- 1.69	146	5.32 +/- 2.58
Finnish ¹	79	7.45 +/- 3.52	31	4.38 +/- 2.22	43	6.44 +/- 3.11
Swedish	307	7.03 +/- 3.31	140	4.07 +/- 2.04	87	6.11 +/- 2.93
Russian ²	50	6.12 +/- 2.96	19	2.90 +/- 1.59	7	NA
TMRCA (ky)	1399	–	618	36.7 +/- 5.6	400	68.4 +/- 13.4

¹From Finnilä *et al.* (2001)

²From Malyarchuk & Derenko (2001) and Loogväli *et al.* (2004)

does not perform quantification and statistical testing of the trends one may intuitively observe. For this purpose we applied the SAMOVA also for analysis of haplogroups. Jointly, these analyses yielded an understanding of the history of the main haplogroups of the region, which allowed a better interpretation of haplogroup frequency patterns across the studied region.

The Y-chromosomal Haplogroups

The frequency distribution and age of haplogroup N3 in our study sample was consistent with the earlier studies (Lahermo *et al.* 1999, Zerjal *et al.* 2001, Tambets *et al.* 2004, Karlsson *et al.* 2006, Rootsi *et al.* 2007). According to the YHRD database, the haplotypes most common in Finland and Karelia were relatively unique, which is not unexpected, since data from most Eurasian populations where N3 is common is not publicly available. It seems evident that the Finns and Karelians share a history regarding haplogroup N3. In the database comparisons, we also observed that N3 may mark a westward diffusion in the north from Finland to Sweden and in the south from the Baltic countries to Poland and Germany.

It has been suggested (Zerjal *et al.* 2001, Roewer *et al.* 2005) that the differences in the haplotype structure of Baltic-speaking Latvians and Lithuanians and Finno-Ugric populations, also observed in our N3 data, imply that the migrations introducing N3 to the region followed a bifurcating pattern, which discourages the idea of a language switch among Latvians and Lithuanians (Laitinen *et al.* 2002). Another factor contributing to the pattern of genetic variation could

be the divergence by linguistic isolation between Baltic and Finnic speakers. However, such a process would affect all haplogroups, which is not supported by genetic distances calculated from the total Y-chromosomal data set – the strong divergence of the Baltic speakers is at least in part specific for N3. A founder effect among the Baltic populations is not any better an explanation, since the diversity among the Baltic speakers is not decreased. Thus, our data supports the idea of two migrations that introduced N3 to the Baltic Sea region (Fig. 5). The haplotype variation in Estonia suggests an admixture of Baltic and Finno-Ugric haplotypes. Furthermore, a bifurcating migration pattern can contribute to the relatively high age of the haplogroup in the Baltic region, since the coalescence age represents the common root of the total variation in the region.

Haplogroup I1a is suggested to have its origins in the Iberian refugium, from where it spread northward and now has its highest frequencies in Northern Europe (Rootsi *et al.* 2004). The haplotype matches to Germany and Poland imply that I1a has arrived to the Nordic countries from the Southern Baltic Sea region, which is historically plausible. The coalescence age of the haplogroup is about 5000 years lower than the age of the earliest archaeological findings from the Northern Baltic Sea region, which suggests a Neolithic arrival. There are two possible migration routes from Central Europe to the Northern Baltic Sea region: an exclusive western route via Sweden, an eastern route via the Baltic states, or via both to Eastern Finland and Karelia (Fig. 5). The surprisingly high diversities of I1a among the eastern Finnish and Baltic populations, and the lack of association between the Western Finns and the Swedes in SAMOVA analysis

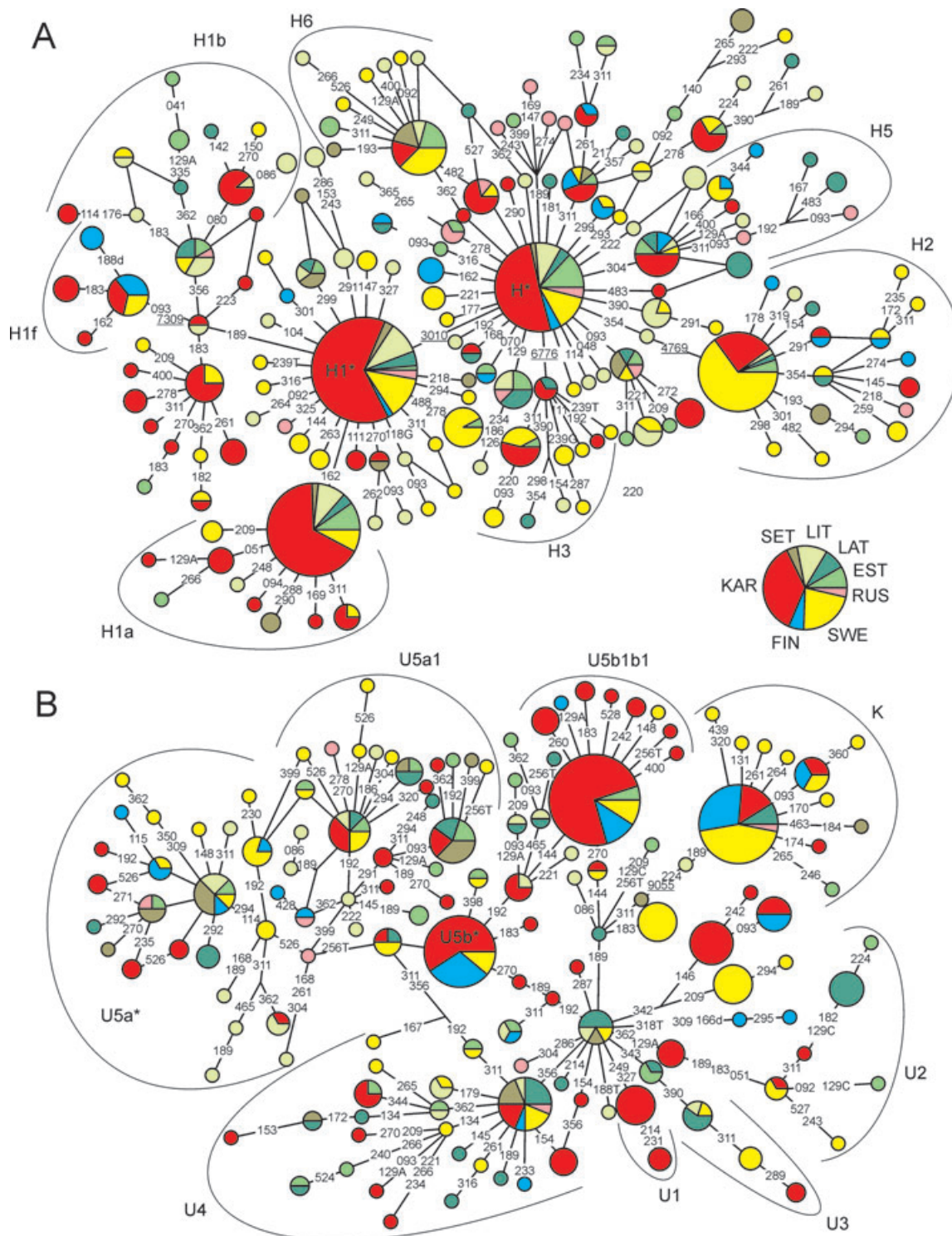


Figure 4 Median-joining networks of the haplotype variation of mitochondrial DNA haplogroups H (A) and U (B). The lengths of the branches correspond to the number of mutations, and the sizes of the nodes denote the number of samples of the corresponding haplotype. The HVS1 mutation sites are given $-16\ 000$, and the coding region polymorphisms are underlined. The bases are given only for the sites where there are more than two alleles; for the others, see Supplementary Table 3. Note that position 16519 is not included in the network. The pie chart denotes the proportions of the different populations in the total sample set of the mtDNA. For the abbreviations, see Figure 2.

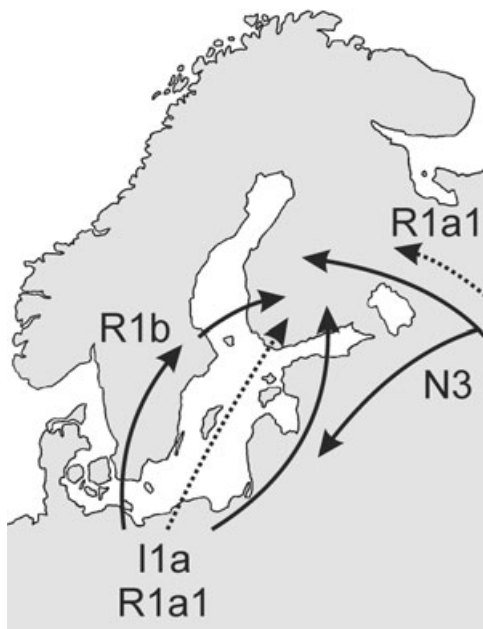


Figure 5 Suggested arrival routes of the most important Y-chromosomal haplogroups in the Baltic area, with the dotted arrows denoting less certain routes.

suggest that I1a has been involved in bifurcating migrations both via Sweden and the Baltic states, and that the presence of the haplogroup in Finland and Karelia is not merely due to Swedish influence. The low frequency of I1a among the Baltic populations may be due to later effects of genetic drift or replacement.

Haplogroup R1a1 is known to be most prevalent in Eastern Europe, and has possibly expanded alongside the Kurgan culture and/or the Indo-European language (Semino *et al.* 2000). The Baltic and Swedish haplotypes had affinities mainly with Germany and Poland in database comparisons, which suggests gene flow from that region to the Western and Eastern coasts of the Baltic Sea. It is plausible that both R1a1 and I1a were carried to the Baltic Sea region via the same Neolithic migrations from Germany/Poland. The higher coalescence age and the starlike network structure of R1a1 are consistent with the probable higher diversity and frequency of R1a1 in the original source population(s), a consequence of the wider geographical distribution of the haplogroup. It is an important observation that in the Baltic Sea region R1a1 is mainly associated to Central European rather than eastern or Russian influence. However, haplotype frequency comparisons (Derenko *et al.* 2006, Willuweit & Roewer 2007) give some indication of Russian gene flow as a partial source of R1a1 in Karelia, which would be plausible given the long period of admixture with Slavs (Fig. 5). However, the Y-chromosomal diversity in Karelia has been heavily affected by drift and

founder effects. Another haplogroup with eastern affinity is I1b (Rootsi *et al.* 2004), whose presence in Karelia and the Baltic states is probably a sign of Russian gene flow.

Haplogroup R1b, the most common haplogroup of Western Europe (Rosser *et al.* 2000, Semino *et al.* 2000), was relatively rare in the northern Baltic Sea region, which, with the exception of Sweden, shows a lack of recent West European influences. R1b has been suggested as having its origins in the Iberian refugia (Semino *et al.* 2000) in a similar manner to I1a (Rootsi *et al.* 2004) but their frequency distribution points to a very different history for R1b and I1a, at least in Northern Europe.

Mitochondrial DNA Haplogroups

The specificity of H1f for the Finnish population has been associated with drift within Finland (Loogväli *et al.* 2004), which is supported by the low number of high frequency haplotypes in the network. Its abundance also in Aunus and Viena Karelia and Ingria provides a strong support for the close historical ties between the Finns and Karelians. The high frequency of haplogroup H2 among the Swedes may be due to sampling bias or local genetic drift, since most of the Swedish H2 samples belong to a single haplotype.

Haplogroups H1, H3, U5b and V have been associated with the expansion from the Iberian refugia after the Ice Age (Torroni *et al.* 1998, 2006, Achilli *et al.* 2004, 2005). Interestingly, H1 and U5b have frequency peaks in Northern Europe in addition to the Iberian peninsula, and our Karelian sample even had a higher H1 frequency than that of the Basques (Achilli *et al.* 2004, 2005, Torroni *et al.* 2006). We suggest three possible scenarios to explain this pattern. First, it is possible but unlikely that genetic drift has consistently increased the frequencies of these haplogroups in the entirety of Northern Europe, though it probably has contributed to the frequency peaks among the Karelians. Another explanation for the high frequencies would be migration from Southern to Northern Europe, possibly via the Atlantic and Baltic coasts, but the archaeological evidence for this is limited. The third alternative scenario would be initially high H1 and U5b frequencies in the entirety of Europe that were partly replaced by other haplogroups in Central Europe due to subsequent migrations that did not affect the north and the southwest. We consider this to be historically and genetically plausible, but further research is needed to analyze this in detail. Haplogroups H3 and V, despite their similar origin in the Iberian refugia, do not follow the same frequency pattern as H1 and U5b, as their frequencies in the Baltic Sea region are barely any higher than in Central Europe. However, the low frequencies of these haplogroups across Europe make reliable comparisons difficult, and even though there is evidence for

a similar historical scheme, they are still distinct haplogroups with possibly different histories.

Haplogroup U is an ancient European haplogroup with an age as great as 55,000 years (Richards *et al.* 2000), and it had a very old coalescence of age also in our analysis. U5b1b1, the so-called “Saami motif”, was very common among the Karelian populations especially in Viena, which, together with the high frequency of D5 and the presence of Z, is a clear sign of shared population history for the Saami and Karelians (Ingman & Gyllenstein 2006). This is also supported by archaeological evidence, since many prehistoric cultural features of the Saami people have arrived from the east via Karelia.

The eastern elements in the mtDNA variation of the Baltic Sea region are intertwined with the Saami influence. Recent studies of the mtDNA variation among the Saami show a link to the Volga-Ural region (Tambets *et al.* 2004, Ingman & Gyllenstein 2006), which is now shown to exist also among the Karelians and, to a lesser degree, among the other populations from the Baltic Sea region as well. Additionally, the presence of U4 in the Eastern Baltic Sea populations may represent eastern influence, since it is typical for the Volga-Ural region (Bermisheva *et al.* 2002). The high diversity of this haplogroup in the Baltic region, observable in the haplotype network, suggests a complex history, and rules genetic drift out as a cause of the high frequency. All in all, these mtDNA haplogroups may be maternal reflections of the eastern influence that can be most clearly observed in the Y-chromosomal haplogroup N3.

Population-based Y-chromosomal and mtDNA Analyses

Genetic distances between the populations are in accordance with linguistic and geographic boundaries in our study and also previous analyses (Zerjal *et al.* 2001) regarding the close proximity of the Baltic countries, and the close association of Finland and Karelia. Admixture analysis of the Estonians supports the importance of geographical location compared to language in determining genetic variation. The Finnish Y-chromosomal variation has been affected by drift, and the grouping of Finns in the paternal MDS plot highlights their bimodal population structure (Kittles *et al.* 1998, 1999, Lahermo *et al.* 1999, Lappalainen *et al.* 2006). The Karelians share the low diversity of the Finns also on the maternal side, especially the Viena Karelians, isolated by their northern geographical location, and the Tver Karelians among whom a founder effect has probably taken place. The diversities of the Balts and the Swedes were in accordance with common European values, and reflect their less isolated geographical position and a probably more stable population size.

Conclusions

The populations of the Baltic Sea region have their strongest roots in Central Europe, which is compatible with archaeological information regarding the arrival of the first inhabitants and several later prehistorical cultures. Additionally, the populations from the eastern side of the Baltic Sea region carry signs of migrations rooted in the east that may be associated to the Finno-Ugric language and/or the Comb Ceramic culture. Less pronounced local admixtures with other populations include Saami influence especially among the Karelians, Russian admixture among them and the Balts, and Central European influence in Sweden. An interesting phenomenon our data has confirmed is the common features between the Iberian peninsula and Northern Europe, observed especially in mtDNA variation, and we hope that future research will shed light on its cause. Furthermore, several Y-chromosomal and mtDNA haplogroups in the Baltic Sea region are of paleolithic origin in Europe. This is consistent with earlier observations (Richards *et al.* 2000) suggesting preservation or even enrichment of the traces of the most ancient European settlement in the northern periphery of the continent. In conclusion, this study also provided strong evidence that the Baltic Sea has been a route not only for trade and cultural exchange but also for population migration.

Acknowledgments

The study was financially supported by the Emil Aaltonen Foundation, the Finnish Cultural Foundation, and the Research Foundation of the University of Helsinki. The Finnish Genome Center provided the genotyping platforms for the STRs and the SNPs by Sequenom[®]. We thank Pertti Sistonen, Tuula Koski and Richard Villems for providing samples, and Ella Granö for technical assistance. Finally, we would like to express our gratitude to the sample donors for their contribution.

References

- Achilli, A., Rengo, C., Battaglia, V., Pala, M., Olivieri, A., Fornarino, S., Magri, C., Scozzari, R., Babudri, N., Santachiara-Benerecetti, A. S., Bandelt, H. J., Semino, O., Torrioni, A. (2005) Saami and Berbers—an unexpected mitochondrial DNA link. *Am. J. Hum. Genet.* **76**, 883–6.
- Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V., Moral, P., Dugoujon, J. M., Roostalu, U., Loogväli, E. L., Kivisild, T., Bandelt, H. J., Richards, M., Villems, R., Santachiara-Benerecetti, A. S., Semino, O., Torrioni, A. (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am. J. Hum. Genet.* **75**, 910–8.
- Alenius, K. (2000) *Viron, Latvian ja Liettuan historia* (The History of Estonia, Latvia and Lithuania). Jyväskylä, Atena.
- Bandelt, H.-J., Forster, P., Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48.

- Bandelt, H.-J., Forster, P., Sykes, B. C., Richards, M. B. (1995) Mitochondrial portraits of human populations. *Genetics* **141**, 743–753.
- Bermisheva, M., Tambets, K., Villems, R., Khusnutdinova, E. (2002) Diversity of mitochondrial DNA haplotypes in ethnic populations of the Volga-Ural region of Russia. *Mol. Biol. (Mosk.)* **36**, 990–1001.
- Bosch, E., Calafell, F., Santos, F. R., Perez-Lezaun, A., Comas, D., Benchemsi, N., Tyler-Smith, C., Bertranpetit, J. (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. *Am. J. Hum. Genet.* **65**, 1623–38.
- Derenko, M., Malyarchuk, B., Denisova, G. A., Wozniak, M., Dambueva, I., Dorzhu, C., Luzina, F., Miscicka-Sliwka, D., Zakharov, I. (2006) Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Hum. Genet.* **118**, 591–604.
- Dupanloup, I. & Bertorelle, G. (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol. Biol. Evol.* **18**, 672–675.
- Dupanloup, I., Schneider, S., Excoffier, L. (2002) A simulated annealing approach to define the genetic structure of populations. *Mol. Ecol.* **11**, 2571–81.
- Finnilä, S., Lehtonen, M. S., Majamaa, K. (2001) Phylogenetic network for European mtDNA. *Am. J. Hum. Genet.* **68**, 1475–84.
- Forster, P., Harding, R., Torroni, A., Bandelt, H. J. (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.* **59**, 935–45.
- Huurre, M. (1990) *9000 vuotta Suomen esihistoriaa (9000 years of Finnish prehistory)*. Keuruu, Otava.
- Ingman, M. & Gyllensten, U. (2006) A recent genetic link between Sami and the Volga-Ural region of Russia. *Eur. J. Hum. Genet.* **15**, 115–20.
- Karlsson, A. O., Wallerstrom, T., Gotheystrom, A., Holmlund, G. (2006) Y-chromosome diversity in Sweden – a long-time perspective. *Eur. J. Hum. Genet.* **14**, 963–70.
- Kittles, R. A., Bergen, A.W., Urbanek, M., Virkkunen, M., Linnoila, M., Goldman, D., Long, J. C. (1999) Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. *Am. J. Phys. Anthropol.* **108**, 381–99.
- Kittles, R. A., Perola, M., Peltonen, L., Bergen, A.W., Aragon, R. A., Virkkunen, M., Linnoila, M., Goldman, D., Long, J. C. (1998) Dual origins of Finns revealed by Y chromosome haplotype variation. *Am. J. Hum. Genet.* **62**, 1171–9.
- Kivisild, T., Tolk, H. V., Parik, J., Wang, Y., Papiha, S. S., Bandelt, H. J., Villems, R. (2002) The emerging limbs and twigs of the East Asian mtDNA tree. *Mol. Biol. Evol.* **19**, 1737–51.
- Lahermo, P., Savontaus, M. L., Sistonen, P., Beres, J., de Knijff, P., Aula, P., Sajantila, A. (1999) Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. *Eur. J. Hum. Genet.* **7**, 447–58.
- Laitinen, V., Lahermo, P., Sistonen, P., Savontaus, M. L. (2002) Y-chromosomal diversity suggests that Baltic males share common Finno-Ugric-speaking forefathers. *Hum. Hered.* **53**, 68–78.
- Lappalainen, T., Koivumaki, S., Salmela, E., Huoponen, K., Sistonen, P., Savontaus, M. L., Lahermo, P. (2006) Regional differences among the Finns: a Y-chromosomal perspective. *Gene* **19**, 207–15.
- Loogväli, E. L., Roostalu, U., Malyarchuk, B. A., Derenko, M. V., Kivisild, T., Metspalu, E., Tambets, K., Reidla, M., Tolk, H. V., Parik, J., Pennarun, E., Laos, S., Lunkina, A., Golubenko, M., Barac, L., Pericic, M., Balanovsky, O. P., Gusar, V., Khusnutdinova, E. K., Stepanov, V., Puzyrev, V., Rudan, P., Balanovska, E. V., Grechanina, E., Richard, C., Moisan, J. P., Chaventre, A., Anagnou, N. P., Pappa, K. I., Michalodimitrakis, E. N., Claustres, M., Golge, M., Mikerezi, I., Usanga, E., Villems, R. (2004) Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol. Biol. Evol.* **21**, 2012–21.
- Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B., Torroni, A. (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **64**, 232–49.
- Malyarchuk, B. A., Derenko, M. V. (2001) Mitochondrial DNA variability in Russians and Ukrainians, implication to the origin of the Eastern Slavs. *Ann. Hum. Genet.* **65**, 63–78.
- Meyer, S., Weiss, G., von Haeseler, A. (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* **152**, 1103–10.
- Nei, M. (1987) *Molecular Evolutionary Genetics*. New York, Columbia University Press.
- R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A., Bandelt, H. J. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **67**, 1251–76.
- Richards, M. B., Macaulay, V. A., Bandelt, H. J., Sykes, B. C. (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann. Hum. Genet.* **62**, 241–60.
- Roewer, L., Croucher, P. J., Willuweit, S., Lu, T. T., Kayser, M., Lessig, R., de Knijff, P., Jobling, M. A., Tyler-Smith, C., Krawczak, M. (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Hum. Genet.* **116**, 279–91.
- Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., Help, H., Bermisheva, M., Kutuev, I., Barac, L., Pericic, M., Balanovsky, O., Pshenichnov, A., Dion, D., Grobei, M., Zhivotovsky, L. A., Battaglia, V., Achilli, A., Al-Zahery, N., Parik, J., King, R., Cinnioglu, C., Khusnutdinova, E., Rudan, P., Balanovska, E., Schefstfahn, W., Simonescu, M., Brehm, A., Goncalves, R., Rosa, A., Moisan, J. P., Chaventre, A., Ferak, V., Furedi, S., Oefner, P. J., Shen, P., Beckman, L., Mikerezi, I., Terzic, R., Primorac, D., Cambon-Thomsen, A., Krumina, A., Torroni, A., Underhill, P. A., Santachiara-Benerecetti, A. S., Villems, R., Semino, O. (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am. J. Hum. Genet.* **75**, 128–37.
- Rootsi, S., Zhivotovsky, L. A., Baldovic, M., Kayser, M., Kutuev, I. A., Khusainova, R., Bermisheva, M. A., Gubina, M., Fedorova, S. A., Ilumae, A. M., Khusnutdinova, E. K., Voevoda, M. I., Osipova, L. P., Stoneking, M., Lin, A. A., Ferak, V., Parik, J., Kivisild, T., Underhill, P. A., Villems, R. (2007) A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur. J. Hum. Genet.* **15**, 204–11.
- Rosser, Z. H., Zerjal, T., Hurler, M. E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W. et al. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.* **67**, 1526–43.

- Schneider, S., Roessler, D., Excoffier, L. (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.
- Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A. S., Cavalli-Sforza, L. L., Underhill, P. A. (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* **10**, 1155–9.
- Siiriäinen, A. (2003) The Stone and Bronze Ages; in Helle, K., Jansson, T. (eds), *The Cambridge History of Scandinavia*. Cambridge, Cambridge University Press.
- Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–62.
- Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogvali, E. L., Tolk, H. V., Reidla, M., Metspalu, E., Pliss, L., Balanovsky, O., Pshenichnov, A., Balanovska, E., Gubina, M., Zhadanov, S., Osipova, L., Damba, L., Voevoda, M., Kutuev, I., Bermisheva, M., Khusnutdinova, E., Gusar, V., Grechanina, E., Parik, J., Pennarun, E., Richard, C., Chaventre, A., Moisan, J. P., Barac, L., Pericic, M., Rudan, P., Terzic, R., Mikerezi, I., Krumina, A., Baumanis, V., Koziel, S., Rickards, O., De Stefano, G. F., Anagnou, N., Pappa, K. I., Michalodimitrakis, E., Ferak, V., Furedi, S., Komel, R., Beckman, L., Villems, R. (2004) The western and eastern roots of the Saami—the story of genetic “outliers” told by mitochondrial DNA and Y chromosomes. *Am. J. Hum. Genet.* **74**, 661–82.
- Torroni, A., Achilli, A., Macaulay, V., Richards, M., Bandelt, H. J. (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet.* **22**, 339–45.
- Torroni, A., Bandelt, H. J., D’Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M. L., Bonne-Tamir, B., Scozzari, R. (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to north-eastern Europe. *Am. J. Hum. Genet.* **62**, 1137–52.
- Willuweit, S., Roewer, L., on behalf of the International Forensic Y Chromosome User Group (2007) Y chromosome haplotype reference database (YHRD): Update. *For. Sci. Int.: Genetics* **1**, 83–87.
- Wilson, Weale & Balding (2003) Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J. R. Stat. Soc. Ser. A*, **166**, 155–188.
- Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339–48.
- Zerjal, T., Beckman, L., Beckman, G., Mikelsaar, A. V., Krumina, A., Kucinskas, V., Hurles, M. E., Tyler-Smith, C. (2001) Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. *Mol. Biol. Evol.* **18**, 1077–87.
- Zhivotovsky, L. A., Underhill, P. A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T., Scozzari, R., Cruciani, F., Destro-Bisol, G., Spedini, G., Chambers, G. K., Herrera, R. J., Yong, K. K., Gresham, D., Tournev, I., Feldman, M. W., Kalaydjieva, L. (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* **74**, 50–61.

Supplementary Material

The following supplementary material is available for this article:

Table S1. The primers, genotyping protocols and haplogroup assignment criteria.

Table S2. The Y-chromosomal haplotype data. For the biallelic polymorphisms corresponding to the assigned haplogroups, see Supplementary Table 1.

Table S3. The data of mitochondrial DNA haplogroups and the polymorphisms in the region 16037–16528. For the assignment of the haplogroups and the corresponding coding region polymorphisms, see Supplementary Table 1.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1469-1809.2007.00429.x>

(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Received: 29 June 2007

Accepted: 14 December 2007