Original Paper

Human Heredity

Hum Hered 2002;53:68–78

Received: August 24, 2001
Revision received: January 10, 2002
Accepted: January 14, 2002

Y-Chromosomal Diversity Suggests that Baltic Males Share Common Finno-Ugric-Speaking Forefathers

Virpi Laitinen[a, b] Päivi Lahermo[c] Pertti Sistonen[d] Marja-Liisa Savontaus[a, b]

[a]Department of Medical Genetics, Biomedical Institute, University of Turku, Turku,
[b]Department of Genetics, University of Turku, Turku, [c]Finnish Genome Center, University of Helsinki, Helsinki, and
[d]Finnish Red Cross Blood Transfusion Service, Helsinki, Finland

Key Words
Y chromosome · SNPs · Haplogroup · Baltic region · Finno-Ugric

Abstract

Objective: To elucidate the genetic relationships between Estonian, Latvian and Lithuanian men by studying Y-chromosomal variation in these people. ***Methods:*** The allelic status of five deep-rooted marker loci (YAP, Tat, M9, 92R7 and *SRY*-1532) was determined for 346 Baltic males. On the basis of single nucleotide polymorphism (SNP) haplotypes, Y chromosomes were divided into six haplogroups, and the Baltic haplogroup distribution compared with that in 7 European reference populations. Haplogroup frequencies, diversities and genetic distances (F_{ST} values) were calculated. The relationships between populations were further illustrated using Mantel test, neighbor-joining tree and principal-component map. ***Results:*** We found the Indo-European-speaking Latvians and Lithuanians to be genetically very similar to the Finno-Ugric-speaking Estonians. When compared to the reference populations, Baltic males were most closely related to the Finno-Ugric-speaking Mari, followed by their Finnish and Slavonic neighbors. ***Conclusions:*** The genetic similarity existing between Estonian, Latvian and Lithuanian men suggests that they originate from the same male founder population. Since the Baltic Y-chromosomal haplogroup distribution more closely resembles that of Finno-Ugric than Indo-European-speaking populations, we propose a hypothesis that Baltic males share a common Finno-Ugric ancestry.

Introduction

Estonia, Latvia and Lithuania are small North European countries which are surrounded by the Baltic Sea in the west, the Gulf of Finland in the north, Russia in the east, Belarus in the southeast and Poland in the south (fig. 1). Although living geographically close to each other, the Baltic people speak languages representing two completely different language groups. Estonian belongs to the Finno-Ugric branch of the Uralic languages, together with e.g. Finnish. Other more distantly related Finno-Ugric languages comprise, e.g., the Saami languages, Mari, Moksha, Komi and Hungarian. With the exception of Basque, most other languages spoken in Europe, including Latvian and Lithuanian, are members of the Indo-

KARGER

Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0001–5652/02/0532–0068$18.50/0

Accessible online at:
www.karger.com/journals/hhe

Dr. Virpi Laitinen
Department of Medical Genetics, Biomedical Institute, University of Turku
Kiinamyllynkatu 10
FIN–20520 Turku (Finland)
Tel. +358 2 333 7456, Fax +358 2 333 7300, E-Mail virlaiti@utu.fi

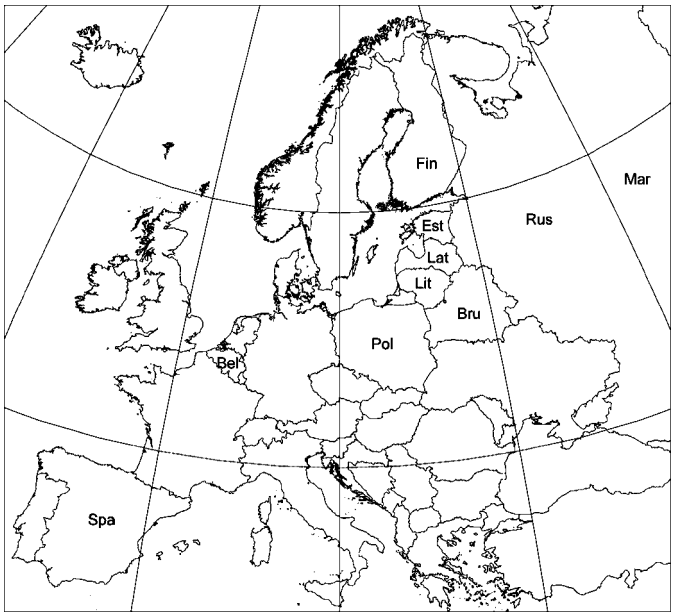

Fig. 1. Populations included in this study. Source map: ESRI Data & Maps 1999. Abbreviated population names: Est = Estonians, Lat = Latvians, Lit = Lithuanians, Fin = Finns, Mar = Mari, Pol = Poles, Rus = Russians, Bru = Belarussians, Bel = Belgians, Spa = The Spanish.

European language group. Therefore, an interesting question arises: are Baltic peoples genetically related to each other or not? Linguistic affiliations would seem to indicate that they are not, whereas geography and the lack of any major geographic barriers do not argue against relatively close relationships amongst them.

The early stages of colonization of the Baltic countries are poorly known. It is possible that the Baltic area had its first inhabitants as early as 35,000–40,000 years ago, at the time when anatomically modern humans spread to Europe [1, 2]. The origin, language or ethnic group of these Paleolithic peoples are unknown, since most, if not all, archaeological evidence was destroyed by the continental glacier that covered the whole Northern Eurasia during the last glacial maximum approximately 23,000 BC [3]. Most likely, these people were hunter-gatherers. Difficult living conditions restricted the size of the population to hundreds, perhaps a few thousands of individuals [4].

The ice began to retreat slowly, and by ~10,000 BC the Baltic region was completely exposed. It is therefore not surprising that the earliest archaeological findings of more permanent settlement date back to only ~9,000 BC [4]. But who were these first permanent settlers of Estonia, Latvia and Lithuania? According to the traditional view, the early Mesolithic populations of the Baltic region were neither Finno-Ugric nor Indo-European, but of unknown origin [4]. The supporters of the more recently formulated continuity theory believe that the Baltic area, perhaps even the whole of northern Europe, was inhabited by Uralic-speaking people for at least the past ~10,000 years [3, 5, K. Wiik, pers. commun.]. These people are suggested to have been the descendants of the hunter-gatherers who lived in the periglacial zone between the Carpathian Mountains and the Volga River during the last glacial maximum [5].

Although the origin of the Mesolithic people continues to be debated, scientists seem to agree to a certain extent on the population history of the Baltic region from the middle Neolithic onwards. It has been suggested that Finno-Ugric tribes arrived in the Baltic region from the east or southeast approximately 4,000–3,000 BC [6]. They merged with the original inhabitants who adopted the Finno-Ugric language together with the Comb ware culture of the newcomers [4]. The members of this new Finno-Ugric-speaking ethnic group are regarded as the ancestors of modern Estonians [6]. The next migration wave came from the south ~1,000 years later. These Central and Southeast European populations spread Corded-ware and Battle-Axe cultures, as well as agriculture to the Baltic region [4]. They spoke a proto-Baltic language, and are considered the predecessors of ancient Prussians as well as present-day Latvians and Lithuanians [7]. It is possible that these Indo-European speakers pushed the Finno-Ugric populations up north, across the Daugava, a river splitting present-day Latvia. In this case, deep-rooted markers, such as single nucleotide polymorphisms (SNPs), should reveal genetic differences between northern and southern parts of the Baltic region. Alternatively, when effectively spreading their culture, the Indo-European-speaking immigrants may also have intermingled with the original Finno-Ugric speakers, who, as a result, adopted the new language [4].

Lively trade between the Baltic people and the neighboring Scandinavian and Central European populations characterized the Bronze Age (1,500–500 BC). It is likely that the exchange of merchandise and cultural skills was accompanied with minor population movements [4]. The Iron Age (500–0 BC) is regarded as the time period when the differentiation of the Baltic tribes was completed [6]. Indo-European-speaking immigrants gradually merged with the original Finno-Ugric-speaking population in Estonia, whereas, in Latvia and Lithuania, the Finno-Ugric and other ethnic components were assimilated into the Indo-European-speaking (Baltic) main population [4].

Anthropologists and linguists agree that Estonians are, in many respects, very similar to the Finns and other Finno-Ugric-speaking populations. Geneticists do not question this view either. A few reports concerning classical markers [8] and mtDNA diversity [9] exist, and they indicate that the closest genetic relatives of the Estonians are their Finno-Ugric-speaking neighbors, such as the Finns and the Karelians. Interestingly, recent studies based on both blood group antigens [10, 11] and Y-chromosomal variation [12–14] suggest that the Indo-European-speaking part of the Baltic region is also genetically similar to the surrounding Finno-Ugric populations. Therefore, it seems that linguistic factors have had less influence on shaping the gene pool of the Baltic area.

Little is known about the genetic relationships between the Baltic people even now. With each genetic system studied (classical markers, mtDNA and Y chromosome), Estonians have been the major focus of attention. Less interest has been paid to Latvians and Lithuanians, who have usually appeared as reference populations consisting of only 30–35 samples [9, 12–14]. Two research articles have concentrated on the within-population diversity of serologically defined ABO and Rh(D) blood groups, as well as mtDNA variation in Lithuania [15, 16], but no comparisons were made between the Lithuanians and their Indo-European or Finno-Ugric-speaking neighbors.

In the present study we further elucidate the genetic relationships between the Estonians, the Latvians and the Lithuanians by examining Y-chromosomal diversity in these people. This is the first time Latvian and Lithuanian populations are characterized to this extent ($n > 100$). The allelic status of five deep-rooted SNP markers was determined for 346 Baltic males, and Y-chromosomal haplogroup distribution compared between all three Baltic populations. The Baltic data were also compared with those from surrounding Finno-Ugric and Indo-European-speaking populations, mainly deriving from an earlier study by Rosser et al. [13].

Subjects and Methods

DNA Samples

Y chromosomes of 346 unrelated, healthy Baltic males were analyzed in this study. Of these, 118 were Estonian, 114 were Latvian, and 114 were Lithuanian. Blood samples were collected during blood donation events organized by local Red Cross Blood Transfusion Services or health care centers. The nationality of the donors was confirmed by the birthplace of their maternal and paternal grandparents. Total DNA was extracted from lymphocytes by standard organic extraction methods.

For comparison, the following 1,078 samples from 7 European reference populations were also included in this study: 537 Finnish, 48 Mari, 112 Polish, 122 Russian, 41 Belarussian, 92 Belgian and 126 Spanish samples. All but Finnish Y chromosome data have been published earlier [13], but they had to be modified (see section 'Haplogroup Formation') to make them correspond to our results, which were obtained with a narrower set of restriction enzymes. The 537 Finnish Y chromosomes will be described in detail elsewhere [Koivumäki et al., in preparation].

SNP Loci and Protocols

In order to elucidate the frequency and distribution of Y-chromosomal haplogroups in the Baltic area, the allelic status of five biallelic loci was determined. Markers known to be polymorphic in Northern Europe were chosen on the basis of earlier studies [13, 17, 18]. YAP insertion [19] was analyzed as described by Hammer and Horai [20]. Length variation of the poly-A tail at the 3′ end of the YAP element was not determined, due to the small number of samples found to carry the insertion (5/346). Four SNPs were typed according to the following protocols: Tat as in Zerjal et al. [21], M9 [22] as in Hurles et al. [17], 92R7 [23] as in Hurles et al. [24] and *SRY*-1532 [25] as in Santos et al. [26]. The LLY22g *Hin*dIII polymorphism was typed in a subset of our samples, using the primers and protocols kindly provided by C. Tyler-Smith and T. Zerjal [pers. commun.]. However, since we observed a perfect correspondence between LLY22g and Tat alleles in our samples (i.e. LLY22g C was always accompanied with Tat T and LLY22g A with Tat C), we omitted the former as the loss of data was likely to be only marginal.

Haplogroup Formation

The term haplogroup (HG) refers to a group of related Y chromosomes, which share the same allelic combinations in the evolutionary old SNP loci but may, and often do, differ with respect to alleles present in the faster-evolving microsatellite loci. The principles of haplogroup formation have been described in previous work [13]. Since we analyzed only five SNP loci, the resolution in our study is limited in so far as some of the haplogroups remained undetected.

The haplogroups were named following the nomenclature system of Jobling et al. [27] and Tyler-Smith [28]. Y chromosomes belonging to haplogroup 1 all contain the following SNP alleles: YAP–, Tat T, M9 G, 92R7 T and *SRY*-1532 G. The allelic 'codes' for the other haplogroups detected in this study are: –TCCG for HG2, –TGTA for HG3, +TCCG for HG4, –CGCG for HG16 and –TGCG for HG26.

To be able to compare the Baltic data with those of other European populations, we combined haplogroups 1 and 22 of the study by Rosser et al. [13] to represent HG1, haplogroups 2 and 9 to represent HG2, haplogroups 4, 21 and 8 to represent HG4, and haplogroups 26 and 12 to represent HG26. The frequencies of haplogroups 3 and 16 were adopted as such. These combinations do not bias the original data, because all of the combined haplogroups have the same allelic status with respect to the five biallelic loci we analyzed. For example, Y chromosomes belonging either to HG4, HG21 or HG8 are all of the type +TCCG (in the YAP, Tat, M9, 92R7 and *SRY*-1532 loci, respectively).

Median-joining networks illustrating the connections between different haplogroups were drawn manually based on the principles of Bandelt et al. [29, 30].

Statistical Analyses

Haplogroup frequencies were determined for each population. Diversities and their standard errors were calculated applying the method described by Nei [31]. Genetic distances were obtained from a population pairwise F_{ST} matrix, created by ARLEQUIN version 2.0 [32]. As a molecular distance method, the sum of squared size differences was used. To evaluate the significance of the genetic distances, the associated p values based on 10,100 permutations were calculated using Amova implemented in the ARLEQUIN program package. Matrix correlation analysis by the Mantel test procedure with 10,000 permutations was also performed by ARLEQUIN.

To further investigate population relationships, principal-component (PC) analysis [33] was carried out. In this method $k = 1,..., n$ for each population, where n stands for the number of populations (here $n = 10$), and vector $x_k = (x_{k1}, x_{k2},..., x_{kN})^T$ is obtained as follows. First, each item x_{ki} corresponds to a separate haplogroup (HGXX), e.g. x_{k1} and x_{k2} correspond to HG01 and HG02, respectively. The value of x_{ki} gives the number of times the population has the haplogroup HGXX. Finally, the vectors x_k are normalized to have an equal number of samples in all populations under study. After obtaining the PCs for samples x_k, the population data can be visualized with two PCs (as shown in fig. 5), and the relations of populations can be examined. The calculations were performed using commercial MATLAB software package (Mathworks Inc.).

Sammon mapping belongs to a class of multidimensional scaling algorithms and was performed on the data as described in Sammon [34]. The GDA program [35] was used to calculate coancestry identity between the populations, which could be illustrated in the form of unweighted pair group method with arithmetic mean (UPGMA) or neighbor-joining (NJ) tree.

Results

Biallelic Marker Loci

The frequencies of individual SNP alleles are presented in table 1.

The Y-chromosome-specific Alu insertion (YAP) was rare in the Baltic people. We found the insertion in only three Estonian men (2.5%), and it was even less frequent in the Indo-European-speaking part of the Baltic region, being present in one Latvian and one Lithuanian sample (0.9%).

The C allele of the Tat polymorphism was present at a relatively high frequency in all three Baltic populations. The distribution of the Tat alleles followed a geographic pattern: the frequency of the C allele increased from north to south. Tat C was present in 33.9% of the Estonians, 42.1% of the Latvians and 43.0% of the Lithuanians.

Also the derived A allele of the *SRY*-1532 locus was common in the Baltic area. The highest incidence was observed in Latvia (39.5%), but the frequency was only slightly lower in the two other Baltic countries. As many as 37.3% of the Estonian and 36.0% of the Lithuanian Y chromosomes were found to carry the *SRY*-1532 A allele.

Table 1. Frequencies (%) of individual SNP alleles in the Baltic populations

SNP	Allele	Estonia	Latvia	Lithuania
YAP	–	97.5	99.1	99.1
	+	2.5	0.9	0.9
Tat	T	66.1	57.9	57.0
	C	33.9	42.1	43.0
M9	C	22.0	8.8	17.5
	G	78.0	91.2	82.5
92R7	C	57.6	50.9	60.5
	T	42.4	49.1	39.5
SRY-1532	A	37.3	39.5	36.0
	G	62.7	60.5	64.0

The G allele of the M9 polymorphism clearly predominated, accounting for 78.0% of the Estonian, 82.5% of the Lithuanian and 91.2% of the Latvian Y chromosomes. The C and T alleles of the 92R7 locus were distributed evenly among the Baltic males: the frequency of the C allele was 50.9% in Latvia, followed by 57.6% in Estonia and 60.5% in Lithuania.

Haplogroup Diversity

Based on the biallelic marker data obtained in this study, Baltic Y chromosomes were divided into specific clusters or haplogroups, as in Helgason et al. [18] and Rosser et al. [13]. With our set of SNPs, a total of seven different haplogroups could be defined. All the 346 Baltic Y chromosomes could be assigned to six haplogroups, which were haplogroups 1–4, 16 and 26. The ancestral haplogroup 7 was absent, as expected, since it has so far been detected only in sub-Saharan Africans [13]. Haplogroup frequencies and diversities for the three Baltic populations are listed in table 2. The Estonians were most diverse, harboring Y chromosomes belonging to all six haplogroups. However, HG26 chromosomes were not detected in the Latvians or the Lithuanians. Haplogroups 16 and 3 were the most predominant in all three populations, accounting for 39.6 and 37.6% of the Baltic Y chromosomes, respectively. The majority of the remaining Y chromosomes belonged either to HG2 (14.7%) or HG1 (6.1%). Five Baltic males were found to represent HG4 (1.4%), and only two Estonians could be placed to HG26 (0.6%).

In general, haplogroup frequencies in the three Baltic populations resembled each other relatively closely (table 2). However, the incidence of HG1 was slightly higher and that of HG2 clearly lower in Latvia than in the two

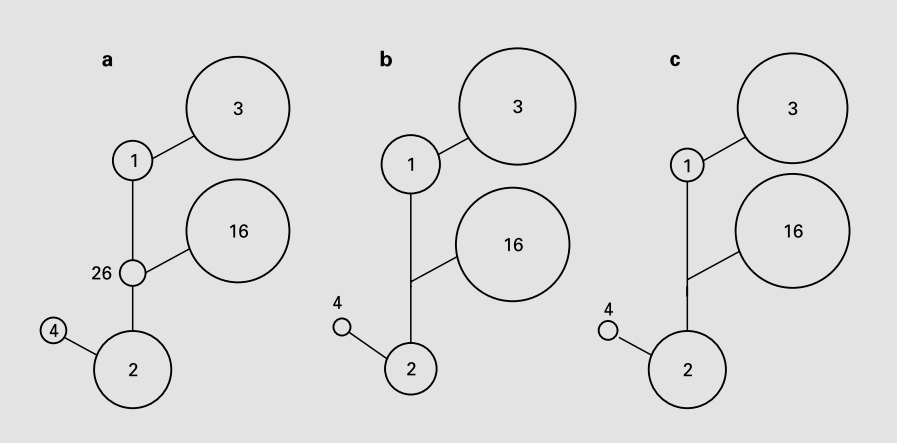

Fig. 2. Median-joining networks representing Y-chromosomal haplogroup distribution in Estonia (a), Latvia (b) and Lithuania (c). The area of the circle is proportional to the frequency of the haplogroup in the population. Haplogroups are identified by numbers. Lines denote mutation events.

Table 2. Y-chromosomal haplogroup frequencies (%) and diversities in the Baltic populations

Haplogroups	Estonia (n = 118)	Latvia (n = 114)	Lithuania (n = 114)	All (n = 346)
HG1	5.1 (6)	9.6 (11)	3.5 (4)	6.1 (21)
HG2	19.5 (23)	7.9 (9)	16.7 (19)	14.7 (51)
HG3	37.3 (44)	39.5 (45)	36.0 (41)	37.6 (130)
HG4	2.5 (3)	0.9 (1)	0.9 (1)	1.4 (5)
HG16	33.9 (40)	42.1 (48)	43.0 (49)	39.6 (137)
HG26	1.7 (2)	0	0	0.6 (2)
HG diversity	0.711 (±0.020)	0.657 (±0.024)	0.663 (±0.022)	

The allelic states of the biallelic loci are as follows: –TGTG for HG1, –TCCG for HG2, –TGTA for HG3, +TCCG for HG4, –CGCG for HG16, and –TGCG for HG26. The order of the biallelic loci is: YAP, Tat, M9, 92R7 and *SRY*-1532. The number of individuals representing each haplogroup is shown in parentheses.

Table 3. F_{ST} distances ($\times$ 100) between the Baltic and reference populations

	Est	Lat	Lit	Fin	Mar	Pol	Rus	Bru	Bel	Spa
Est	–									
Lat	1	–								
Lit	0	0	–							
Fin	15	19	12	–						
Mar	0	0	0	14	–					
Pol	11	13	15	40	15	–				
Rus	3	6	6	27	5	3	–			
Bru	7	14	12	30	12	7	2	–		
Bel	16	21	21	34	19	17	13	12	–	
Spa	19	23	23	37	21	19	15	15	0	–

Nonsignificant distances (based on 10,100 permutations) at the 5% level are underlined. Population names are abbreviated as in figure 1.

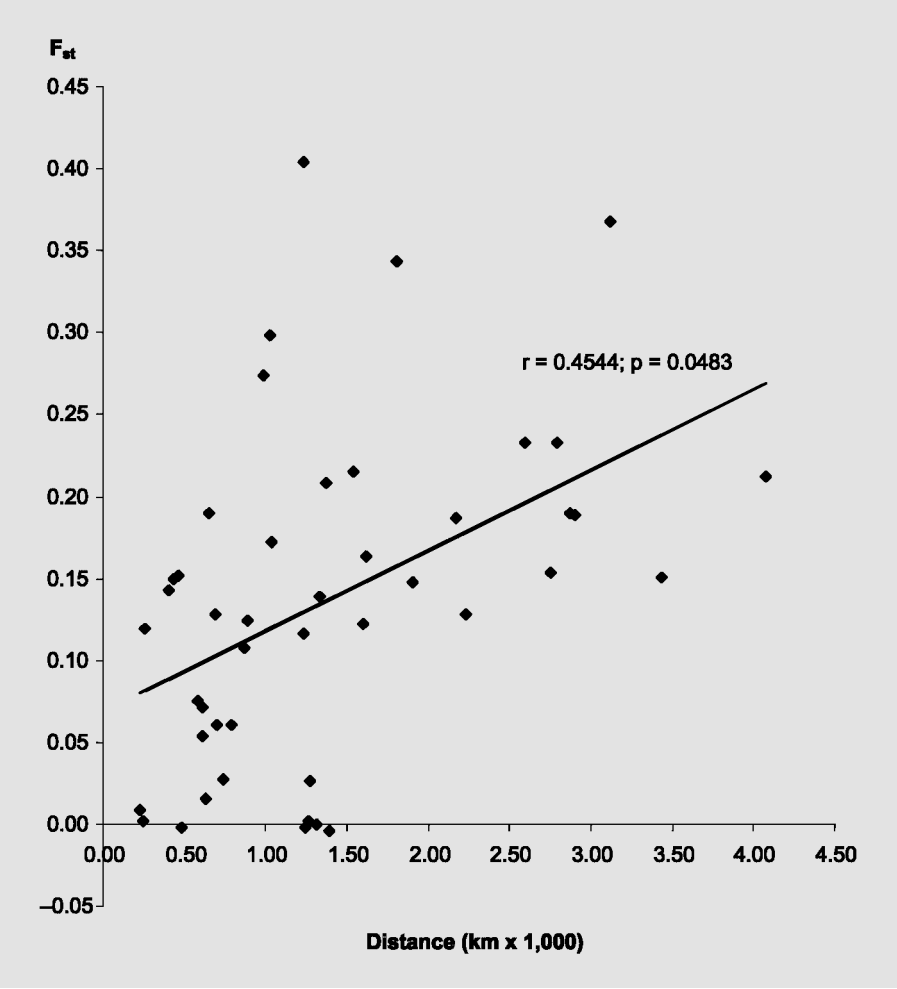

Fig. 3. Correlation analysis of genetic distance (F_{ST} values calculated on the basis of SNP haplotypes) and geographic distance (km) for the Baltic people and the seven European reference populations.

other Baltic countries. Also, HG16 chromosomes were less common in Estonia than in Latvia or Lithuania. To illustrate the proportions of different haplogroups in these populations, median-joining networks were constructed (fig. 2).

Genetic versus Geographic Distances

In addition to the Baltic people, reference populations representing different regions and different linguistic groups of Europe were included in the computation of genetic distances (table 3). Shortest distances were observed among Estonians, Latvians, Lithuanians and the Finno-Ugric-speaking Mari, indicating close genetic relationships between these four North European populations. The Baltic people were also fairly closely related to Russians from the Moscow area, followed by the Belarussians, the Poles and the Finns. The Belgians and the Spanish were clearly distinct from the Baltic populations, as well as from the Mari and the Finns.

Using the Mantel test, population pairwise F_{ST} values were compared to geographic distances (km) between all of the 10 populations. The results of this test revealed a slightly significant positive correlation ($r = 0.4544$, $p = 0.0483$) (fig. 3).

In UPGMA (not shown) and NJ trees (fig. 4), Estonians, Latvians and Lithuanians consistently appeared in the same branch as the Mari. PC analysis confirmed the observation that the Baltic populations grouped together, with the Mari as their closest relative. The PC distances of the Baltic populations to their Finnish and Slavonic neighbors were approximately similar, the Indo-European-speaking populations being the most distant group (fig. 5). Sammon mapping gave essentially the same result (data not shown).

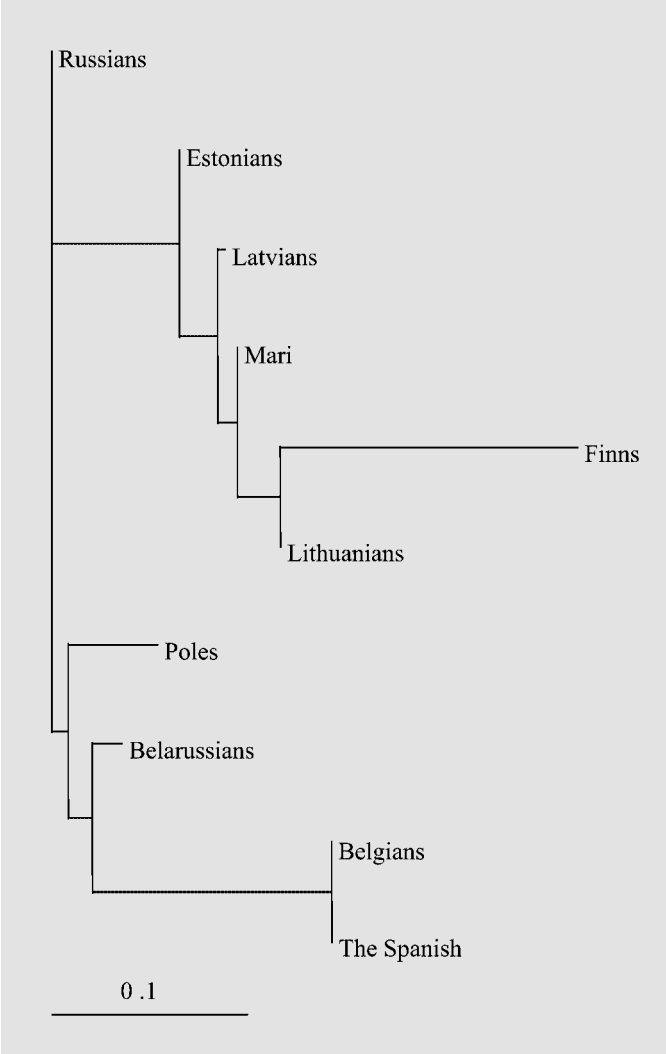

Fig. 4. An NJ tree illustrating the phylogenetic relationships between the Balts and the seven European reference populations.

Discussion

The haplogroup-defining mutations used in this study represent unique or nearly unique events in human evolution, and are believed to have happened only once, with the exception of *SRY*-1532, which has undergone a recurrent mutation [28]. The ages of all these mutations vary from thousands of years for Tat T→C [21, 36], 92R7 C→T [17, 37] and *SRY*-1532 G→A [36, 38] to even tens of thousands of years for Alu insertion [39] and M9 C→G [22]. Therefore, they can serve as indicators of evolutionary events that took place in human history several thousands of years ago.

Alu Insertion

As expected, the frequency of Alu insertion was low in Estonia (table 1). The YAP+ type is rarely seen in Finno-Ugric-speaking populations, with the exception of the ethnic Csángós and other Hungarians, who carry the insertion in 38 and 18% of their Y chromosomes, respectively [12]. Instead, the insertion has been reported to be present in approximately 10% of Indo-European-speaking men [19], to whom Latvians and Lithuanians belong. Although recent studies [18, 40] have suggested that the insertion is less common in Northern Europe, it has nonetheless been detected in 2–3% of men inhabiting the British Isles, Scandinavia and Russia [18]. Therefore, we expected the insertion to be more common in Latvia and Lithuania than what we observed (0.9%). When the frequency is compared with that of neighboring populations, the differences are too small to allow far-reaching conclusions to be made. But it is rather tempting to speculate that the low frequency of Alu insertion in the Indo-European-speaking part of the Baltic region could be indicative of a possible Finno-Ugric influence on these people.

Tat C and HG16 Chromosomes

Y chromosomes carrying the C allele of the Tat polymorphism belong to haplogroup 16. Most C alleles are found among speakers of Altaic or Uralic languages. The highest frequency of the C allele has been detected in the Siberian Yakut (86%), followed by Buryats (58%) who live in Mongolia [21]. The C allele is also predominant among the Finns (61%), the Karelians (40%) and the Mari (33%), who all belong to the speakers of Finno-Ugric languages [12]. By contrast, most Indo-European speakers harbor the T allele in their Y chromosomes. The C allele is relatively rare among the Swedes (7%) and the Slovakians (3%), and completely absent in British, Italian and Bulgarian males [13, 21].

It has recently been reported that in the Baltic area, linguistic boundaries have had little, if any impact on the distribution of the Tat C allele [12–14]. This fits well with our data, since we found the C allele to be common not only among the Finno-Ugric-speaking Estonians, but also among the Indo-European-speaking Latvians and Lithuanians (table 1). Indeed, the C allele, which is usually regarded as an indicator of Finno-Ugric origin, was more frequent in Latvians (42%) and Lithuanians (43%) than in Estonians (34%). When compared to neighboring Indo-European-speaking populations, a dramatic decrease in the frequency of HG16 chromosomes could be seen. Only 14% of the Russians (Moscow area), 4% of the Poles and 2% of the Belarussians carry the Tat C allele

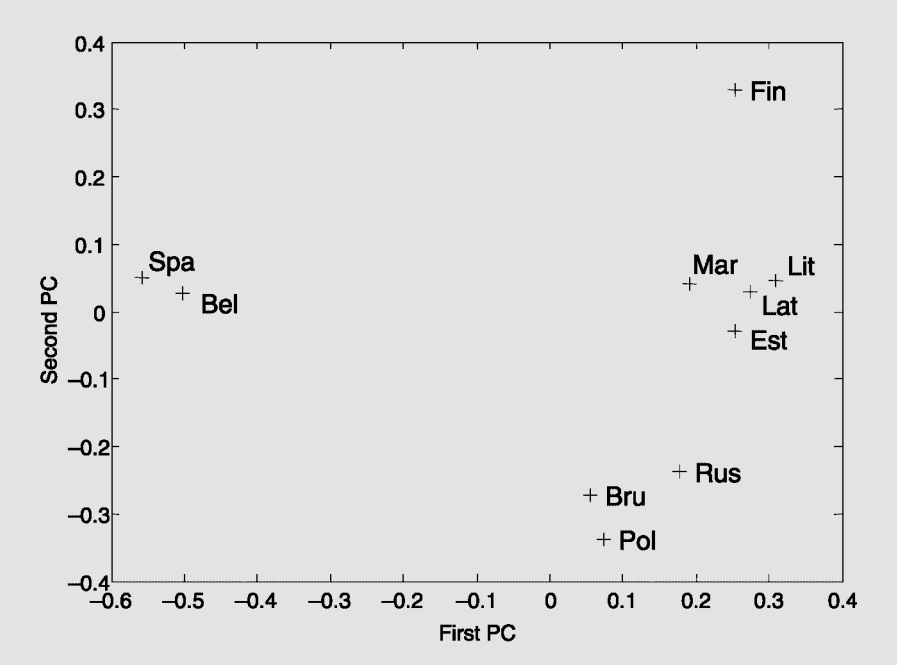

Fig. 5. A two-dimensional PC map constructed from Y-chromosomal haplogroup frequency data (table 1). The first PC retains 65.7% of the variance, the second 25.3% (cumulatively 91.0% of the variance). Population names are abbreviated as in figure 1.

[13]. It is unlikely that HG16 chromosomes would have been enriched in the Baltic area by chance, as a result of similar and simultaneous evolutionary processes, such as bottlenecks or genetic drift. Rather, the high frequency of Tat C allele in all three Baltic countries could reflect the genetic composition of the male founder population, suggesting common eastern – possibly Finno-Ugric – roots for these people. This interpretation is supported by a recent, comprehensive study of Y-chromosomal diversity in Europe [13], where HG16 chromosomes were proposed to represent population movements from Eastern Europe or Northern Asia to the Baltic Sea.

SRY-1532 A and HG3 Chromosomes

Y chromosomes carrying the A allele in the *SRY*-1532 locus are assigned to haplogroup 3, when they harbor the derived states of M9 and 92R7 loci as well (G and T alleles, respectively). HG3 chromosomes are typically found in Central and Eastern Europe, such as Poland (54%), Russia (47%) and Belarus (39%) [13], but they are also widespread and fairly common in Asia [38]. In the Baltic area, more than one third of Y chromosomes belonged to haplogroup 3 (table 2), which again speaks for a possible eastern influence in the males of these three populations.

Genetic Similarity between the Balts and the Mari

The Y-chromosomal haplogroup distribution was similar in Estonia, Latvia and Lithuania (table 2, fig. 2). More than three quarters of the Baltic Y chromosomes could be placed into only two haplogroups (haplogroups 3 and 16) and three of the haplogroups encompassed ~92% of the Y chromosomes in the Baltic area (haplogroups 2, 3 and 16). However, small differences in the frequencies of individual haplogroups between populations could be seen. These differences may be due to founder effects, genetic drift, immigration and/or gene flow, but these evolutionary phenomena have most likely taken place after the male founder population had settled in the Baltic area several thousands of years ago. When compared to the seven reference populations, Baltic haplogroup distribution resembled closely only that in the Mari, a Finno-Ugric-speaking population inhabiting an area between the Volga and the Ural. Also the genetic distances between these four populations were effectively zero (table 3). Instead, genetic distances between the Baltic people and one other Finno-Ugric reference population included in this study, the Finns, were clearly longer. This can be explained by the Finnish haplogroup distribution [unpubl. data], which was rather different from the Baltic one. There was, however, one important feature that the Finnish and the Baltic males had in common: a high frequency of eastern HG16 chromosomes.

The neighboring Indo-European-speaking populations, the Russians, the Poles and the Belarussians, were found to be more distantly related to the Baltic men than the Mari (table 3, fig. 5). The most striking differences were the low frequency of HG16 chromosomes, and, on the other hand, the higher incidence of HG2 chromosomes in these three reference populations. The Belgian and the Spanish Y chromosomes represented Central European type, and differed significantly from the Baltic ones. In Belgium and Spain, HG1 chromosomes were clearly predominant (with frequencies of 63 and 68%, respectively), whereas only 6.1% of the Baltic Y chromosomes belonged to HG1. Furthermore, no HG16 chromosomes were found in the Belgian and Spanish samples [13].

As could be expected on the basis of similar haplogroup distribution, various statistical analyses failed to reveal any significant differences between the three Baltic populations. In the PC (fig. 5) and Sammon maps (not shown), Baltic people clustered very closely together with the Mari, but remained clearly separated from the rest of the populations. The same phenomenon was again repeated in the phylogenetic NJ (fig. 4) and UPGMA trees (not shown), obtained from the coancestry coefficients. Regardless of the statistical evaluation method used, the outcome was always the same: the males of Estonia, Latvia and Lithuania were genetically closely related, and showed a marked degree of similarity to the Finno-Ugric-speaking Mari. In our opinion, this favors the hypothesis that the Baltic males may descend from the same Finno-Ugric-speaking founder population. This view is further supported by the positive correlation between geographic and genetic distances obtained in the Mantel test (fig. 3). Alternative models, such as migration and gene flow, fail to explain why the geographically distant Mari are genetically so close to the Baltic males.

Our findings challenge those recently reported by Zerjal et al. [14], who suggested that the genetic history of Latvian and Lithuanian Y chromosomes is distinct from that of the Uralic speakers. Their conclusion was based on e.g. multidimensional scaling of haplogroup frequencies, which showed that the Latvians and the Lithuanians clustered together but remained separated from their Indo-European- and Finno-Ugric-speaking neighbors (including the Estonians). The discrepancy in the results may be attributed to several reasons. First, Zerjal et al. [14] analyzed a smaller number of samples. Second, they typed as many as 10 binary marker loci; thus, considerably more information was available for each subject. Third, samples characterized in these two studies may have been collected from different regions of the Baltic countries, and may therefore reflect possible regional differences in allele frequencies. And last, it should be borne in mind that the Y chromosome effectively represents a single locus that is strongly influenced by stochastic variation which even alone, but especially together with any of the above-mentioned reasons, may have caused the stated discrepancies.

The Language of the Ancestors?

It is logical to assume that the members of the same language family are also genetically more closely related to each other than to the speakers of unrelated languages. However, recent studies [13, 24] have suggested that language might not always be the major factor in determining relatedness, but geographical vicinity can also play an important role – especially in the absence of barriers, such as high mountains or large water systems. This clearly seems to be the case in the Baltic region. Published data based on the careful analysis of both genetic [10, 11, 14] and craniological [5, 6] material reveal marked similarities between the three Baltic populations, indicating that at least in this geographical territory the linguistic unit does not correspond to the genetic one.

Our results suggest that the Baltic peoples may have a common origin. Since Estonians speak a Finno-Ugric language, and Latvians and Lithuanians belong to the speakers of Indo-European languages, one or more of these populations must have changed languages during the course of history. While the Finno-Ugric-speaking Mari closely resembled the three Baltic populations, none of the Indo-European-speaking reference populations showed close genetic relationships to the Baltic people. Therefore, it seems more likely that the Estonians have always spoken a Finno-Ugric language, whereas the Latvians and the Lithuanians may have replaced their original Finno-Ugric language with an Indo-European one. And indeed, the Latvians still stress the first syllable of each word, which is typical for Finno-Ugric languages [41]. Since language can be adopted similarly to any other cultural skill, without the need for genetic admixture, major migrations are not necessary [5].

Recent archaeological findings [3] suggest that Finno-Ugric-speaking tribes have inhabited the shores of the Baltic Sea continuously for several millennia. This view, known as the continuity theory, is strongly supported by the evidence obtained from physical anthropology studies. The facial features and craniometric measures of the Baltic-Finns are not significantly different from those of other European populations, favoring the idea that Finno-

Ugric speakers have lived in the Baltic area for a very long time [5]. Similar results have been reported earlier by Cesnys [6], who proposed that the early inhabitants of the Baltic region had arrived to the coast of the Baltic Sea from the east, possibly as early as 8,000 BC. Assuming that Estonia, Latvia and Lithuania have indeed been settled by Uralic-speaking tribes since the early Mesolithic, the most logical explanation for our Y-chromosomal data would be a common Finno-Ugric ancestry for the males of these three Baltic populations.

Acknowledgements

We wish to express our thanks to all our blood sample donors. Dr. Richard Villems from the University of Tartu and Estonian Biocentre is acknowledged for contributing the Estonian DNA samples. MSc Satu Koivumäki from the Department of Genetics of the University of Turku kindly provided the Finnish SNP data. MSc Marjaana Salonen and BSc Reetta Paananen are thanked for their excellent technical assistance in sample analysis at the Department of Medical Genetics of the University of Turku. We are also indebted to Dr. Leena Laurila from the Department of Geography of the University of Turku for editing the map in figure 1. This work was financially supported by the Academy of Finland (grant No. 38826) and Turun Suomalainen Yliopistoseura ry.

References

1 Cann RL, Stoneking M, Wilson AC: Mitochondrial DNA and human evolution. Nature 1987; 325:31–36.

2 Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ: Y chromosome sequence variation and the history of human populations. Nat Genet 2000;26:358–361.

3 Nuñez M: Uusi katsaus Suomen asuttamismalliin; in Julku K (ed): Itämerensuomi – eurooppalainen maa. Jyväskylä, Atena Kustannus Oy, 1997, pp 47–62.

4 Alenius K: Viron, Latvian ja Liettuan historia. Jyväskylä, Atena Kustannus Oy, 2000.

5 Niskanen M: Itämerensuomalaisten alkuperä fyysisen antropologian näkökulmasta; in Julku K (ed): Itämerensuomi – eurooppalainen maa. Jyväskylä, Atena Kustannus Oy, 1997, pp 104–120.

6 Cesnys G: Anthropological roots of Lithuanians. Science, Arts & Lithuania 1991;1:4–10.

7 Niitemaa V, Hovi K: Baltian historia, ed 2, rev. Helsinki, Kustannusosakeyhtiö Tammi, 1991.

8 Beckman L, Sikström C, Mikelsaar A-V, Krumina A, Ambrasiene D, Kucinskas V, Beckman G: Transferrin variants as markers of migrations and admixture between populations in the Baltic Sea region. Hum Hered 1998;48: 185–191.

9 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, DiRienzo A, Pääbo S: Genes and languages in Europe: An analysis of mitochondrial lineages. Genome Res 1995;5:42–52.

10 Sistonen P, Virtaranta-Knowles K, Denisova R, Kucinskas V, Ambrasiene D, Beckman L: The LW^b blood group as a marker of prehistoric Baltic migrations and admixture. Hum Hered 1999;49:154–158.

11 Heapost L: A population genetic characterization of Estonians. Anthrop Anz 2000;58:137–154.

12 Lahermo P, Savontaus M-L, Sistonen P, Béres J, de Knijff P, Aula P, Sajantila A: Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. Eur J Hum Genet 1999;7:447–458.

13 Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Côrte-Real HBSM, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar A-V, Mitchell RJ, Nafa K, Nicholson J, Norby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA: Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 2000;67:1526–1543.

14 Zerjal T, Beckman L, Beckman G, Mikelsaar A-V, Krumina A, Kucinskas V, Hurles ME, Tyler-Smith C: Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. Mol Biol Evol 2001;18: 1077–1087.

15 Kucinskas V, Radikas J, Rasmuson M: Genetic diversity in the Lithuanian rural population as illustrated by variation in the ABO and Rh(D) blood groups. Hum Hered 1994;44:344–349.

16 Kucinskas V: Human mitochondrial DNA variation in Lithuania. Anthrop Anz 1994;52: 289–295.

17 Hurles ME, Irven C, Nicholson J, Taylor PG, Santos FR, Loughlin J, Jobling MA, Sykes BC: European Y-chromosomal lineages in Polynesians: A contrast to the population structure revealed by mtDNA. Am J Hum Genet 1998; 63:1793–1806.

18 Helgason A, Sigurdardóttir S, Nicholson J, Sykes B, Hill EW, Bradley DG, Bosnes V, Gulcher JR, Ward R, Stefánsson K: Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. Am J Hum Genet 2000;67: 697–717.

19 Hammer MF: A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 1994;11:749–761.

20 Hammer MF, Horai S: Y chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 1995;56:951–962.

21 Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C: Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 1997;60:1174–1183.

22 Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ: Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 1997;7:996–1005.

23 Mathias N, Bayés M, Tyler-Smith C: Highly informative compound haplotypes for the human Y chromosome. Hum Mol Genet 1994;3: 115–123.

24 Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Pérez-Lezaun A, Bosch E, Shlumukova M, Cambon-Thomsen A, McElreavey K, López de Munain A, Röhl A, Wilson IJ, Singh L, Pandya A, Santos FR, Tyler-Smith C, Jobling MA: Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. Am J Hum Genet 1999;65:1437–1448.

25 Whitfield LS, Sulston JE, Goodfellow PN: Sequence variation of the human Y chromosome. Nature 1995;378:379–380.

26 Santos FR, Pandya A, Tyler-Smith C, Pena SDJ, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ: The central Siberian origin for native American Y chromosomes. Am J Hum Genet 1999;64:619–628.
27 Jobling MA, Pandya A, Tyler-Smith C: The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 1997;110:118–124.
28 Tyler-Smith C: Y-chromosomal DNA markers; in Papiha S, Deka R, Chakraborty R (eds): Genomic diversity: Applications in human population genetics. New York, Plenum Press, 1999, pp 65–73.
29 Bandelt H-J, Forster P, Sykes BC, Richards MB: Mitochondrial portraits of human populations using median networks. Genetics 1995; 141:743–753.
30 Bandelt H-J, Macaulay V, Richards M: Median networks: Speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. Mol Phyl Evol 2000;16: 8–28.
31 Nei M: Molecular evolutionary genetics. New York, Columbia University Press, 1987.
32 Schneider S, Roessli D, Excoffier L: ARLEQUIN: A software for population genetic data analysis. Ver 2.0. Genetics and Biometry Laboratory, University of Geneva, 2000.
33 Cavalli-Sforza LL, Menozzi P, Piazza A: The history and geography of human genes. Princeton, Princeton University Press, 1994.
34 Sammon JW Jr: A nonlinear mapping for data structure analysis. IEEE Trans Comput 1969;5: 401–409.
35 Lewis PO, Zaykin D: Genetic Data Analysis: Computer program for the analysis of allelic data. Ver 1.0 (d16c), 2001. Free program distributed by the authors over the Internet from http://lewis.eeb.uconn.edu/lewishome/software.html.
36 Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF: Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. Am J Hum Genet 1999; 64:817–831.
37 de Knijff P: Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am J Hum Genet 2000;67:1055–1061.
38 Zerjal T, Pandya A, Santos FR, Adhikari R, Tarazona E, Kayser M, Evgrafov O, Singh L, Thangaraj K, Destro-Bisol G, Thomas MG, Qamar R, Mehdi SQ, Rosser ZH, Hurles ME, Jobling MA, Tyler-Smith C: The use of Y-chromosomal DNA variation to investigate population history: Recent male spread in Asia and Europe; in Papiha S, Deka R, Chakraborty R (eds): Genomic diversity: Applications in human population genetics. New York, Plenum Press, 1999, pp 91–101.
39 Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL: Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 1998;15:427–441.
40 Raitio M, Lindroos K, Laukkanen M, Pastinen T, Sistonen P, Sajantila A, Syvänen A-C: Y-chromosomal SNPs in Finno-Ugric-speaking populations analyzed by minisequencing on microarrays. Genome Res 2001;11:471–482.
41 Mela M: Latvian historia, ed 3. Helsinki, Yliopistopaino, 1997.