Y Chromosome and Mitochondrial DNA Variation in Lithuanians

D. Kasperavičiūtė[1], V. Kučinskas[1] and M. Stoneking[2]

[1]Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Santariskiu str. 2, LT-08661 Vilnius, Lithuania, [2]Max Planck Institute for Evolutionary Anthropology; Deutscher Platz 6; D-04103 Leipzig; Germany.

Summary

The genetic composition of the Lithuanian population was investigated by analysing mitochondrial DNA hypervariable region 1, RFLP polymorphisms and Y chromosomal biallelic and STR markers in six ethnolinguistic groups of Lithuanians, to address questions about the origin and genetic structure of the present day population. There were no significant genetic differences among ethnolinguistic groups, and an analysis of molecular variance confirmed the homogeneity of the Lithuanian population. MtDNA diversity revealed that Lithuanians are close to both Slavic (Indo-European) and Finno-Ugric speaking populations of Northern and Eastern Europe. Y-chromosome SNP haplogroup analysis showed Lithuanians to be closest to Latvians and Estonians. Significant differences between Lithuanian and Estonian Y chromosome STR haplotypes suggested that these populations have had different demographic histories. We suggest that the observed pattern of Y chromosome diversity in Lithuanians may be explained by a population bottleneck associated with Indo-European contact. Different Y chromosome STR distributions in Lithuanians and Estonians might be explained by different origins or, alternatively, be the result of some period of isolation and genetic drift after the population split.

Introduction

The territory encompassed today by Lithuania was settled relatively late, as the land became inhabitable only about 12,000 BP, after the last glaciation. Archaeological data showed that the first people to settle were late Palaeolithic hunter-gatherers, belonging to the Swiderian and Baltic Magdalenian cultures (Rimantienė, 1996). Anthropological findings regarding Mesolithic and Neolithic inhabitants of present day Lithuania are very limited (Butrimas et al. 1985) and there is a degree of uncertainty concerning the processes of neolithization, Indo-European dispersal and formation of the Baltic tribes. The Neolithic (6,500–3,500 BP) is a period of intensive cultural development in Lithuania as ceramics, cattle breeding and farming first appeared (Rimantienė, 1996). Linguistic and archaeological studies suggest that the formation of the Balts, who gave rise to the present day Lithuanians and Latvians, took place in the fifth millennium BP. Studies of river and lake names show that, in ancient times, the Baltic languages spread throughout a large territory to the east of the Baltic Sea, from the Vistula in the west, through the entire Upper and middle Dnepr river basin, to the Upper Reaches of the Volga, the Oka and the Moscow river in the east, and up to and across the Pripyat in the south (Vanagas, 1987). This area roughly corresponds to the area in which, during the late Neolithic, three Corded Ware/Boat Axe cultures – Baltic Coastal, Middle Dnepr and Fatyanovo – were spread.

Archaeological evidence suggests that the first Baltic culture in the territory of Lithuania was the Baltic Coastal culture, which was formed in the late Neolithic

*Correspondence to: D. Kasperavičiūtė Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Santariskiu str. 2, LT-08661 Vilnius, Lithuania. E-mail: dalia.kasperaviciute@gf.vu.lt Tel.:+370-5-2365197; Fax:+370-5-2365199

through the interaction of autochthonic (Nemunas and Narva) and Indo-European (Globular Amphora and Corded Ware) cultures (Gimbutas, 1963). This view is also supported by craniometrical anthropological data (Česnys, 2001), though the data are rather limited. Even though it is commonly agreed that the culture of corded ware ceramics had a direct impact on the creation of the first Baltic culture, it is not known whether it influenced autochthonic people more through cultural exchange, or involved large-scale immigration of Indo-Europeans.

Another question concerns the impact of Finno-Ugrian populations on Balts. Traditionally, Finno-Ugrians are associated with comb-pit ceramics; however, as with Baltic cultures, no strong evidence for relating archaeological and ethnical (linguistic) groups exists. The examination of archaeological monuments suggests that Finno-Ugrians probably had at most a minor impact on the Lithuanian population (Rimantienė, 1996). Around 30 hydronames of Finno-Ugrian origin are known throughout Lithuania (Vanagas, 1987; Zinkevičius, 1998), and it is thought that they mark trading roads near the rivers where small groups of people established camps in the mid-fifth millennium BP. The influence of Finno-Ugrians on Balts has also not been detected in analyses of the rare transferrin variants in blood serum in populations of the Baltic Sea region, since the genetic markers of Finno-Ugric influence (alleles TF*DCHI and TF*DFIN) are not found in the Lithuanian and Latvian populations (Beckman *et al.* 1998). However, analysis of Y chromosome biallelic markers revealed surprising similarities of Lithuanians and Latvians to the Finno-Ugric Estonians and Mari (Laitinen *et al.* 2002), which led these authors to the conclusion that Baltic and Finno-Ugrian males share common forefathers.

The large territory inhabited by Balts in ∼3000 BP was at that time covered by virtually impenetrable forests, and was located far from major migration and trade routes. Thus the Balts appear to have lived for a long time in relative isolation. In the middle of the first millennium AD, or somewhat earlier, Slavic tribes began to force their way into the territory inhabited by the Balts. The ancient Baltic territory was split geographically into the western portion – the historic Prussian,

Latvian and Lithuanian lands – and the remaining Baltic tribes in the east, which were eventually assimilated by the Slavs (Zinkevičius, 1998).

Lithuanians are first mentioned in historical documents at the beginning of the 11th century. At that time the name Lithuania was probably not used to denote the entire present-day Lithuanian region, but only a portion comprising the territory between the Neris, Nemunas and Merkys rivers. To the north Lithuanian lands bordered with the lands inhabited by Curonian, Semigalian, Selonian and Lettigalian tribes, the fusion of which gave rise to the Latvian nation, whereas to the south Lithuania bordered with the old Yotvingian lands. The land of Žemaičiai reached to the north and to the south west the early Lithuanians were contiguous with the Prussian tribes. At that time a very intensive consolidation of the related Baltic tribes took place. This contributed greatly to the formation of the present-day Lithuanian dialects. After the formation of the Lithuanian state in the 13th century, the movement of people inside the country was very limited. From the 16th till the end of the 19th century approximately 80% of the population consisted of peasants, who could not move from the feudal domain to which they belonged because of serfdom. This resulted in further linguistic differentiation among different regions of Lithuania. Six dialectal groups are distinguished in present day Lithuanian: three groups of Aukštaitish (west, south and east) and three groups of Žemaitish (north, west and south) (Figure 1).

Figure 1 Lithuanian ethno-linguistic groups.

Thus the contemporary population of Lithuania, the subject of the present study, is composed of a complex mixture of former Baltic tribes, with potentially varying influences from Finno-Ugric and Slavic sources, which could have resulted in genetic heterogeneity within Lithuania. Previous studies of the genetic variation of Lithuanians, based on blood groups and mtDNA RFLP variation (Kučinskas, 1994; 2001) revealed some internal variation, notably the differences between North Žemaičiai and South Aukštaičiai; however, the results were inconclusive. Our previous study of mtDNA HV1 variation in 120 individuals did not show regional differences between two major groups of Lithuanians – Aukštaičiai and Žemaičiai (Kasperavičiūtė & Kučinskas, 2002). Since the linguistic differentiation of Lithuanians is a recent process, Y chromosome STR polymorphisms, which are rapidly mutating markers, might be more informative for exploring genetic heterogeneity within Lithuania, as well as for dissection of relationships with closely related populations. Moreover, since Lithuanians and Latvians live on the border of Indo-European and Finno-Ugric speaking populations, the genetic relationships of these populations might clarify the contacts between these groups in Northern Europe. A previous study of Y chromosome markers (Laitinen *et al.* 2002) suggested a common origin of these populations, while an earlier study of Y-STR variation (Zerjal *et al.* 2001) revealed a genetic boundary between Estonians and Latvians, determined mainly by differences in Y-STR variation associated with haplogroup N3 Y chromosomes. However, this study included only 30-40 individuals from each population and only five STRs were analysed.

In the present study we further elucidate the origins of Lithuanians by examining more mitochondrial DNA and Y chromosome markers in a much larger sample, in comparison with surrounding Indo-European and Finno-Ugric speaking reference populations, to address the following questions:

- Are Lithuanian ethnolinguistic groups genetically differentiated? If so, since dialectal differentiation of Lithuanians was influenced by relationships and contacts with different Baltic tribes and other neighbouring populations, is this evident in the gene pool of the present day Lithuanians?

- How are Lithuanians genetically related to other European, in particular neighbouring Indo-European and Finno-Ugric speaking, populations? Can these relationships be reconciled with historical data on the origins of Lithuanians?

- Do paternal and maternal histories of Lithuanians differ?

- Do genetic relationships correlate with geographic relationships?

Materials and Methods

DNA Samples

Peripheral blood samples were collected from unrelated individuals from six ethno-linguistic groups of Lithuania (Figure 1). Informed consent and information about birthplace, parents and grandparents were obtained from all donors. Genomic DNA was extracted using a standard salting out procedure (Miller *et al.* 1988). For comparative analyses 91 Estonians and 46 Latvians were also sampled from throughout these countries.

Mitochondrial DNA Genotyping

Mitochondrial DNA variation was analysed in 180 Lithuanian samples (30 from each of the six ethnolinguistic groups), including 120 samples described previously (Kasperavičiūtė & Kučinskas, 2002). Primers L15996 and 16401 (Vigilant *et al.* 1989) were used for amplifying the first hypervariable segment (HV1) of the mtDNA control region. These primers were used for direct sequencing of both strands of HV1 using the ABI PRISM Big Dye Terminators cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) on an ABI 310 automated DNA sequencer (Applied Biosystems, Foster City, CA, USA) following the protocol recommended by the supplier. The status of positions 00073, 7028 and 14766 was determined by restriction enzyme digestion with *Alw*44I, *Alu*I and *Mse*I respectively in all samples. The primers L29 and H408 (Vigilant *et al.* 1989) were used for amplification of the HV2 segment containing position 00073, while the typing for positions 7028 and 14766 was performed as described in Torroni *et al.* (1997). On the basis of specific nucleotide substitutions in HV1 and positions 00073, 7028 and 14766, the

sequences were classified into specific European haplogroups, which were confirmed by additional restriction fragment length polymorphism typing according to Macaulay *et al.* (1999), using primer pairs and conditions as in Torroni *et al.* (1997). The HV1 haplotype and RFLP data are available as supplementary material and from the authors; HV1 sequences were also deposited in the HvrBase database (Handt *et al.* 1998).

For phylogenetic analysis, the available published data on HV1 sequences in European, Near East and Caucasus populations were retrieved from the HvrBase database (Handt *et al.* 1998); in addition HV1 sequences from 436 poles and 200 russians (Malyarchuk *et al.* 2002) were included.

Y Chromosome Genotyping

Y chromosome variation was analysed in 196 male samples from Lithuania: 40 from North Aukštaičiai, 34 from South Aukštaičiai, 32 from West Aukštaičiai, 26 from North Žemaičiai, 34 from South Žemaičiai and 30 from West Žemaičiai. Five binary markers known to be polymorphic in Northern Europe were typed using previously described procedures: Tat as in Zerjal *et al.* (1997), YAP (Hammer *et al.* 1994) as in Hammer & Horai (1995), M9 (Underhill *et al.* 1997) as in Kayser *et al.* (2000a), SRY-1532 (Whitfield *et al.* 1995) as in Santos *et al.* (1999), and 92R7 (Mathias *et al.* 1994) as in Hurles *et al.* (1999). Haplogroups were defined by biallelic markers according to the nomenclature proposed by the Y Chromosome Consortium (2002). Y chromosomes containing the following alleles: YAP–, Tat T, M9 G, 92R7 T and SRY-1532 G belong to haplogroup P*(xR1a). The allelic 'codes' for the other haplogroups detected in this study are: − TCCG for haplogroup BR*(xDE,JR), − TGTA for haplogroup R1a, + TCCG for haplogroup DE, − CGCG for haplogroup N3 and − TGCG for haplogroup K*(xN3,P). The correspondence between this nomenclature and the former nomenclature of Jobling *et al.* (1997) and Tyler-Smith (1999) is as follows: P*(xR1a) = Hg1, BR*(xDE,JR) = Hg2/9, R1a = Hg3, DE = Hg4/8/21, N3 = Hg16, K*(xN3,P) = Hg12/26.

The Y chromosome STR loci DYS19 (or DYS394), DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 and DYS385 were analyzed using genotyping protocols and allelic nomenclature described by Kayser *et al.* (1997) and Redd *et al.* (1997) on ABI310 and ABI377 genetic analyzers. The DYS385I and DYS385II loci of haplogroup N3 chromosomes were typed separately, using protocols by Kittler *et al* (2003). Y chromosome haplotypes are available as supplementary material.

Statistical Analysis

The basic parameters of molecular diversity, genetic distances and population genetic structure (including analysis of molecular variance) were calculated using the computer program Arlequin 2.0 (Schneider *et al.* 2000). Multidimensional scaling (MDS) was performed by means of STATISTICA, based on F_{ST} distances between pairs of populations. Lithuanians were compared with European populations for which both mtDNA HV1 and Y chromosome SNP data are available. The correlation between genetic distances based on mtDNA and Y chromosome variation, and between genetic and geographic distances between populations, was evaluated by the Mantel test with 10000 permutations using the Arlequin 2.0 program (Schneider *et al.* 2000).

The phylogenetic relationships between Y chromosome STR haplotypes within haplogroups, as defined by the biallelic loci, were reconstructed in the form of median joining networks (Bandelt *et al.* 1999), with use of the program Network 3.1.1.1 (www.fluxus-engineering.com). Because the repeat length of DYS389II includes DYS389I, the latter was subtracted from DYS389II to avoid double-counting variation at DYS389I. For network calculation, locus-specific weights were used according to the observed mutation rates for the Y-STR loci used here (Kayser *et al.* 2000b), so that loci with the highest mutation rates were given the lowest weights (ratio of DYS393: DYS392: DYS385I: DYS385II: DYS19: DYS389I: DYS389II: DYS391: DYS390 = 10: 10: 7: 7: 5: 5: 2: 2: 1).

Bayesian-based coalescence analyses of Y-STR haplotype data were performed by means of the program BATWING (University of Aberdeen Department of Mathematical Sciences). The principles of the Markov-chain Monte Carlo-based inference method implemented in this program have been described elsewhere (Wilson & Balding, 1998). We chose a two-phase

population model, in which, in the past, the population was of constant size N, followed by a period of exponential growth up to the present. For the initial population size we used a lognormal (6,1) prior distribution, with mode 148, median 403, and mean 665, which represents a rather small initial founder population. The prior population growth rate was an exponential distribution with mean 0.01, which covers the simple constant-population-size model, as well as reasonable growth rates for human populations. The length of the growth period (in units of N times generation time) also had an exponential prior, with mean 1. For the Y-STR mutation rate, first we assigned gamma-distributed prior distributions to the mutation rates of the STR loci, adjusted to the corresponding estimates in Kayser *et al.* (2000b). For each locus we chose a gamma distribution, such that the mode was equal to the mutation rate estimate and the SD was inversely related to the number of meioses investigated by Kayser *et al.* (2000b). Analyses were repeated using a more conservative phylogenetic mutation rate estimate (Zhivotovsky *et al.* 2004) as follows: a prior gamma distribution with mode and SD equal to the Y-STR mutation rate estimate (6.9×10^{-4} per 25 years) and its SD (5.7×10^{-4}) from Zhivotovsky *et al.* (2004) was used. To produce the results, we sampled 20,000 times from the Markov chain after discarding the first 2,000 samples. We checked the robustness of our results using a variety of prior distributions for population parameters. The results were stable under different population parameters priors (data not shown), but depended on the mutation rate estimate (discussed below).

Results

MtDNA Variation

Sequences of the mtDNA HV1 region comprising nucleotide positions 16024-16400 (Anderson *et al.* 1981) were determined for 180 Lithuanians. The data were checked for the presence of systematic sequencing errors (phantom mutations) using the phylogenetic method described by Bandelt *et al.* (2002); no phantom mutations were found (analysis not shown). Analyses were restricted to 356 bp (nucleotide positions 16028-16383) of HV1 for the purpose of comparing the sequences reported here with published data. We detected 76 polymorphic sites and 95 distinct haplotypes among 180 Lithuanians. The proportion of transitions was 92.4%. The length polymorphisms of A and C stretches in region 16180-16193 (triggered by the 16189 T/C substitution) were disregarded in the analyses.

Table 1 reports the mtDNA HV1 sequence diversity indices for all Lithuanian ethnolinguistic groups. The gene diversity (0.947-0.984) and the mean number of pairwise differences (3.61-4.98) estimates are within the range found in European populations (Comas *et al.* 1997, Helgason *et al.* 2000). The distributions of observed number of differences between pairs of sequences (mismatch distributions) in Lithuanians were unimodal and approximately bell shaped. The raggedness indices

Table 1 Diversity and demographic parameters deduced from mtDNA HV1 sequences in Lithuania

Ethnolinguistic group	Sample size (n)	Number of haplotypes	Gene diversity $+/-$SD	Mean number of pairwise nucleotide differences $+/-$SD	Tajima's D value	Raggedness index
Aukštaičiai:						
East Aukštaičiai	30	23	0.982$+/-$0.013	3.79$+/-$1.96	-2.035	0.030
South Aukštaičiai	30	23	0.949$+/-$0.033	4.12$+/-$2.11	-2.000	0.016
West Aukštaičiai	30	24	0.984$+/-$0.013	4.98$+/-$2.49	-1.766	0.011
Aukštaičiai total	90	56	0.972$+/-$0.010	4.29$+/-$2.14	-2.072	0.010
Žemaičiai:						
North Žemaičiai	30	24	0.979$+/-$0.016	4.85$+/-$2.43	-1.647	0.012
South Žemaičiai	30	22	0.947$+/-$0.033	3.61$+/-$1.89	-2.289	0.030
West Žemaičiai	30	24	0.984$+/-$0.013	4.77$+/-$2.40	-1.621	0.011
Žemaičiai total	90	58	0.970$+/-$0.011	4.51$+/-$2.24	-2.056	0.009
Lithuanians total	180	95	0.971$+/-$0.008	4.41$+/-$2.19	-2.051	0.009

Table 2 AMOVA results in the Lithuanian population. Lithuanians were classified into two main groups, Aukštaičiai and Žemaičiai (between group and among population within groups values were not significant, p>0.05 based on 10 000 permutations)

		Percentage of variation	
Grouping	Source of variation	mtDNA HV1 sequences[a]	Y chromosomal STRs[b]
Individual groups	Between groups	0.42	− 0.15
	Within groups	99.58	100.15
Aukštaičiai–Žemaičiai	Between groups	0.32	− 0.83
	Among populations within groups	0.23	0.34
	Within populations	99.45	100.48

a distance method – number of pairwise differences
b distance method – sum of squared size difference between Y-STR haplotypes

varied from 0.01 to 0.03. The bell shaped mismatch distributions together with raggedness indices less than 0.05 are interpreted as signs of prehistoric population expansions (Harpending *et al.* 1993). This was also reinforced by Tajima's (1989) D-statistic, which was significantly negative in all groups (Table 1).

The population structure of Lithuanians was investigated by calculating F_{ST} distances based on HV1 sequences, and by the analysis of molecular variance (AMOVA) method (Excoffier *et al.* 1992). F_{ST} distances between ethnolinguistic groups varied from 0 to 0.028, but were not significantly different from zero. The absence of phylogeographic structure of mtDNA HV1 variation in Lithuanian was further confirmed by AMOVA analysis (Table 2). Both when all six Lithuanian

ethnolinguistic groups were treated as a single group, and when they were grouped into the two main groups of Aukštaičiai and Žemaičiai (based on geographic and linguistic relationships), almost all (∼99.5%) HV1 variation fell within ethnolinguistic groups.

In order to investigate relationships between Lithuanians and other European populations, multidimensional scaling (MDS) based on F_{ST} distances between populations was performed (Figure 2). Overall, European populations are very closely related based on HV1 variation. In the MDS plot Lithuanians lie close to Slavs (Poles and Russians), in between Finno-Ugrian populations of northern Europe and Indo-European populations of Western Europe. However, significant (p<0.01) F_{ST} distances were obtained with only a few
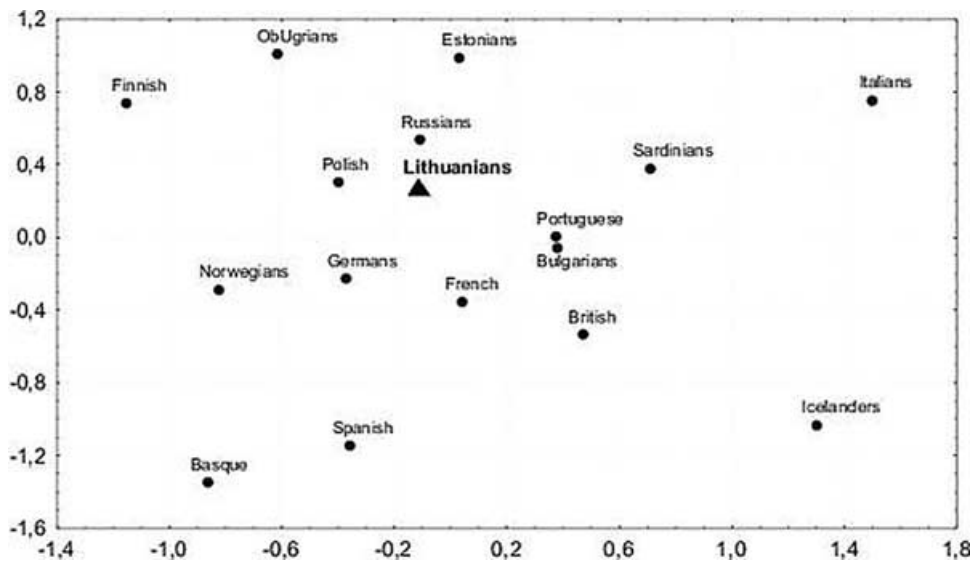

Figure 2 Multidimensional scaling plot of F_{ST} distances between European populations based on mtDNA HV1 variation (stress value 0.153).

Table 3 Frequencies of mtDNA haplogroups in the Lithuanian population

Haplo-group. subhaplo group	Haplogroup frequency (%)								
	East Aukštaičiai (N = 30)	South Aukštaičiai (N = 30)	West Aukštaičiai (N = 30)	North Žemaičiai (N = 30)	South Žemaičiai (N = 30)	West Žemaičiai (N = 30)	Total Aukštaičiai (N = 90)	Total Žemaičiai (N = 90)	Total Lithuanian (N = 180)
H	33.3	26.7	40.0	33.3	46.7	20.0	33.3	33.3	33.3
H1	0	0	0	10.0	0	0	0	3.3	1.7
H3	0	6.7	6.7	10.0	0	3.3	4.4	4.4	4.4
H4	3.3	6.7	0	0	6.7	6.7	3.3	4.4	3.9
H5	3.3	0	0	0	0	3.3	1.1	1.1	1.1
H8	3.3	0	0	0	0	6.7	1.1	2.2	1.7
V	13.3	0	3.3	6.7	3.3	3.3	5.6	4.4	5.0
HV	3.3	3.3	3.3	0	0	0	3.3	0	1.7
preV	3.3	3.3	0	0	0	0	2.2	0	1.1
U	0	3.3	3.3	0	6.7	3.3	2.2	3.3	2.8
K	0	6.7	3.3	0	3.3	0	3.3	1.1	2.2
U3	3.3	3.3	0	0	3.3	0	2.2	1.1	1.7
U4	10.0	3.3	0	0	3.3	13.3	4.4	5.6	5.0
U5a	0	0	6.7	3.3	6.7	0	2.2	3.3	2.8
U5a1	0	3.3	0	3.3	0	0	1.1	1.1	1.1
U5b	3.3	3.3	3.3	6.7	3.3	0	3.3	3.3	3.3
U5b1	3.3	3.3	0	0	0	0	2.2	0	1.1
U5	0	0	0	0	3.3	0	0	1.1	0.6
J	6.7	6.7	3.3	6.7	0	6.7	5.6	4.4	5.0
J1	0	0	0	3.3	0	0	0	1.1	0.6
J1b1	0	0	0	0	3.3	10.0	0	4.4	2.2
T	3.3	10.0	3.3	10.0	3.3	13.3	5.6	8.9	7.2
T1	6.7	3.3	6.7	0	0	0	5.6	0	2.8
I	0	0	10.0	6.7	3.3	3.3	3.3	4.4	3.9
W	0	3.3	0	0	0	3.3	1.1	1.1	1.1
X	0	3.3	0	0	0	0	1.1	0	0.6
Others	0	0	6.7	0	3.3	3.3	2.2	2.2	2.2

geographically distant populations: Italians, Spanish, Basque and Icelanders. To examine the correlation between F_{ST} distances and geographical distances a Mantel test was performed, which revealed a weak (r = 0.41) but significant (p<0.01) correlation.

In the analyses of shared sequences among populations, 32.2% of Lithuanian HV1 sequences were found to be private (not previously detected in other populations). However, no specific combinations of haplotypes or their subclusters that clearly distinguish Lithuanians from neighbouring European populations were found.

Classification of mtDNA sequences revealed the presence of all major European haplogroups in Lithuania, and these haplogroups accounted for 97% of all sequences. Table 3 reports the frequencies of mtDNA haplogroups in Lithuanians. The most frequent haplogroup, comprising almost half of the sequences, is H,

which is also the most frequent haplogroup in Europe and the Near East. However, the specific lineage characterized by 16304C-16311C mutations, which was proposed to mark the Slavonic migrations from Central to East Europe (Malyarchuk & Derenko, 2001), was not found among Lithuanians. Genetic distances between pairs of European populations, based on haplogroup frequencies (adapted from analyses of Helgason *et al.* 2001), were computed using the Cavalli-Sforza method (Cavalli-Sforza & Edwards, 1967). An MDS plot based on these distances revealed a similar position of Lithuanians in Europe as seen in the MDS plot based on HV1 variation (not shown).

Y Chromosome Variation

The biallelic Y chromosome loci used in this study partition Lithuanian Y chromosomes into six haplogroups.

Table 4 Y chromosomal haplogroup frequencies and haplogroup diversities in Lithuanians and some European populations

Population	Sample size (n)	Haplogroup frequecy%						Haplogroup diversity $+/-$ SD
		P* (xR1a)	BR* (xDE,JR)	R1a	DE	N3	K* (xN3,P)	
Aukštaičiai:								
East aukštaičiai	40	0.0	17.5	35.0	2.5	45.0	0.0	$0.660+/-0.039$
South aukštaičiai	34	2.9	2.9	61.8	2.9	29.4	0.0	$0.546+/-0.071$
West aukštaičiai	32	6.3	9.4	40.6	6.3	34.4	3.1	$0.722+/-0.052$
Aukštaičiai total	106	2.8	10.4	45.3	3.8	36.8	0.9	$0.653+/-0.027$
Žemaičiai:								
North žemaičiai	26	15.4	3.9	42.3	0.0	38.5	0.0	$0.674+/-0.050$
South žemaičiai	34	5.9	8.8	50.0	0.0	35.3	0.0	$0.633+/-0.053$
West žemaičiai	30	3.3	16.7	40.0	3.3	36.7	0.0	$0.699+/-0.046$
Žemaičiai total	90	7.8	10.0	44.4	1.1	36.7	0.0	$0.659+/-0.029$
Lithuanians total	196	5.1	10.2	44.9	2.6	36.7	0.5	$0.653+/-0.020$
Latvians[a,b,c]	148	10.8	8.8	39.9	0.7	39.9	0.0	$0.667+/-0.021$
Estonians[a,b,c]	325	7.4	16.9	30.8	2.8	35.7	6.5	$0.741+/-0.012$
Polish[b]	112	17.9	20.6	54.5	1.8	4.5	0.9	$0.633+/-0.037$
Belarusians[b]	41	9.8	36.6	39.0	9.8	2.4	2.4	$0.711+/-0.043$
Rusians[b]	122	6.6	21.3	46.7	6.6	13.9	4.9	$0.712+/-0.031$
Finnish[b]	57	1.8	22.8	10.5	1.8	61.4	1.8	$0.569+/-0.060$

a data from Laitinen *et al.* 2002.
b data from Rosser *et al.* 2000.
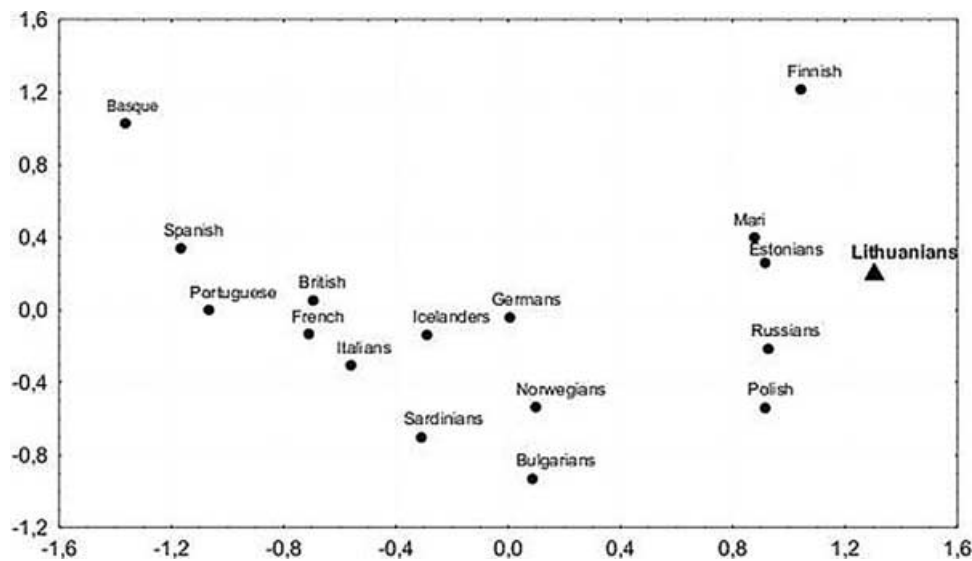c data from this study is not included.

Figure 3 MDS Multidimensional scaling plot of F_{ST} distances between European populations based on Y chromosome biallelic markers (stress value 0.068).

Table 4 reports the frequencies of these six haplogroups in Lithuania and some European reference populations. Two major haplogroups in Lithuanian males are haplogroup R1a and haplogroup N3, comprising 45% and 37%, respectively, of all Y-chromosomes. As noted pre-

viously (Zerjal *et al.* 2001, Laitinen *et al.* 2002), the high frequency of haplogroup N3 places Lithuanians and Latvians closer to Finno–Ugric speaking groups than to Indo–European groups. This is also seen in the MDS plot based on F_{ST} distances (Figure 3). In the MDS

plot Lithuanians are close to Finno–Ugric speaking populations, but (as in the MDS plot based on mtDNA variation) also close to Slavic populations – Russian and Polish. All these populations are clearly different from the remaining European populations. However, only five binary markers, which are known to be polymorphic in Northern Europe, were used to analyze Lithuanian Y chromosomes. This might have resulted in artificially lower diversity in populations of Central and Southern Europe, and in smaller genetic distances between them.

The correlation between F_{ST} distances, based on Y chromosome haplogroup frequencies and geographic distances between populations, was significant (Mantel test, r = 0.45, p<0.001). The correlation between genetic distances based on mtDNA HV1 variation and Y chromosome binary markers was not significant (r = 0.32, p>0.01).

Since linguistic and historical studies suggest that Lithuanian dialectal differentiation is relatively recent, having occurred during the last millennium (Zinkevičius, 1998), fast mutating markers might be more suitable to evaluate genetic differentiation and relationships between ethnolinguistic groups within Lithuania. Therefore we typed nine Y chromosome STR loci in all samples, and constructed compound haplotypes by combining the allelic status of both the binary markers and the STR loci. We detected 123 Y chromosome compound haplotypes among 196 Lithuanian males. Overall, gene diversity based on 9 STR loci was $0.985 +/- 0.004$. Pairwise genetic distances based on STR variation (R_{ST}) within Lithuania were not significantly different from zero (data not shown), thus, similarly to mtDNA variation, there is no genetic differentiation among Lithuanian ethnolinguistic groups. The extreme homogeneity of Lithuanian Y chromosomes was confirmed by analysis of molecular variance (AMOVA), which revealed that all Y chromosome STR variation in Lithuania is due to variation within ethnolinguistic groups (Table 2).

STR diversity analysis within each binary haplogroup revealed that the highest diversity was among haplogroup P^*(xR1a) (n = 10), haplogroup BR^*(xDE,JR) (n = 20) and haplogroup DE (n = 5), in which all STR haplotypes were different (h = 1). This is not surprising, especially for haplogroup BR^*(xDE,JR), which is

known to be a heterogeneous set of Y chromosomes. The gene diversity of haplogroup R1a Y chromosomes was $0.984 +/- 0.005$, where 56 distinct haplotypes were detected and the most frequent of them (16-13-17-25-11-11-13-11,14) was found in 7 individuals, comprising 8% of HgR1a Y chromosomes, and 3.6% of all Y chromosomes in Lithuanian males. Phylogenetic analysis of HgR1a Y chromosomes resulted in a very reticulated median-joining network (not shown).

The lowest gene diversity (h = $0.915 +/- 0.023$) was detected among HgN3 chromosomes. The most frequent haplotype (15-14-16-23-11-14-14-11-13) comprises 25% of HgN3 and 9.2% of all Lithuanian Y chromosomes. This haplotype, together with all one-step mutation neighbours, comprises 61.1% of all HgN3 Y chromosomes (and 22.4% of all Lithuanian Y chromosomes). A median-joining network of Lithuanian HgN3 chromosomes demonstrated that this is also the central haplotype. To determine if this haplotype is specific for Lithuanians, we searched the Ystr database for European populations (http://ystr.charite.de). Surprisingly, only 9 matches among the sample of 13986 European minimal haplotypes were found: 2 in Lithuania (Vilnius), 2 in Latvia (Riga), 4 in Germany and 1 in Norway. Even though the Ystr database does not contain information about binary markers, so that matches might be due to homoplastic mutations on different SNP backgrounds, finding no matches among 399 individuals from Finland and 133 from Estonia (Tallinn) in the database suggests that Lithuanian HgN3 chromosomes might be different from those in Finno-Ugric populations. To further investigate this, we typed 91 Estonian and 46 Latvian males for the Tat marker (which delineates HgN3), and subsequently typed those Y-chromosomes with the TatC alleles for the STR loci. We found 30 (33.3%) and 17 (39.5%) HgN3 Y chromosomes in Estonians and Latvians respectively. One Lithuanian, one Latvian and two Estonian HgN3 Y chromosomes with a duplication of

Table 5 R_{ST} distances and their p values based on 10 000 permutations between East Baltic populations based on HgN3 Y chromosomal STR diversity

	Lithuanians	Latvians	Estonians
Lithuanians	–	<0.00001	<0.00001
Latvians	0.243	–	<0.00001
Estonians	0.114	0.263	–

DYS385II were detected, but excluded from phylogenetic analysis since it was impossible to determine which allele is ancestral and corresponds to DYS385II of other Y chromosomes. The most frequent Lithuanian haplotype (25% HgN3 chromosomes) was detected in only 2 Estonians (6.7%) and not in any Latvian. Lithuanians and Latvians also differed from Estonians for DYS19 alleles: 15 repeats at DYS19 were found in the majority of Lithuanians (94.4% HgN3 chromosomes) and Latvians (88.2%), but only in 40% of Estonians, where the 14 repeat allele was more common (60%), consistent with previous results (Zerjal *et al.* 2001).

Gene diversity and average variance were both lower in Lithuanian ($0.912 +/- 0.023$ and 0.159) and Latvian ($0.917 +/- 0.049$ and 0.158) HgN3 Y chromosomes than in Estonians ($0.943 +/- 0.032$ and 0.196). Pairwise population comparisons (R_{ST} values) of HgN3 chromosomes revealed significant differences among populations (Table 5). In the median-joining network haplotypes also tend to cluster according to populations (Figure 4).

The lower gene diversity and nearly star-like median joining network of Lithuanian HgN3 chromosomes suggests a bottleneck involving HgN3

chromosomes in Lithuania; therefore, a demographic analysis using a Bayesian-based coalescence approach was performed (Table 6). A signal of population growth was detected both in Lithuanians and Estonians (the Latvian sample was excluded from this analysis because of the small sample size), and also for the combined dataset, corrected for differences in sample sizes. The dates of the start of the population growth depended greatly on the mutation rate estimate used for calculations, and varied from ~ 1000 years (using the mutation rate obtained from father-son pairs (Kayser *et al.* 2000b)) to ~ 7-8000 years (using the phylogenetic mutation rate estimate (Zhivotovsky *et al.* 2004)). The true date probably lies in between these estimates. Interestingly, similar dates were obtained for HgR1a Lithuanian chromosomes and all Y chromosomes, indicating that population growth was characteristic for the whole population, while a reduction of diversity is seen only among HgN3 chromosomes.

Discussion

Historically, two main ethnolinguistic groups of Lithuanians, Aukštaičiai and Žemaičiai, developed over a long
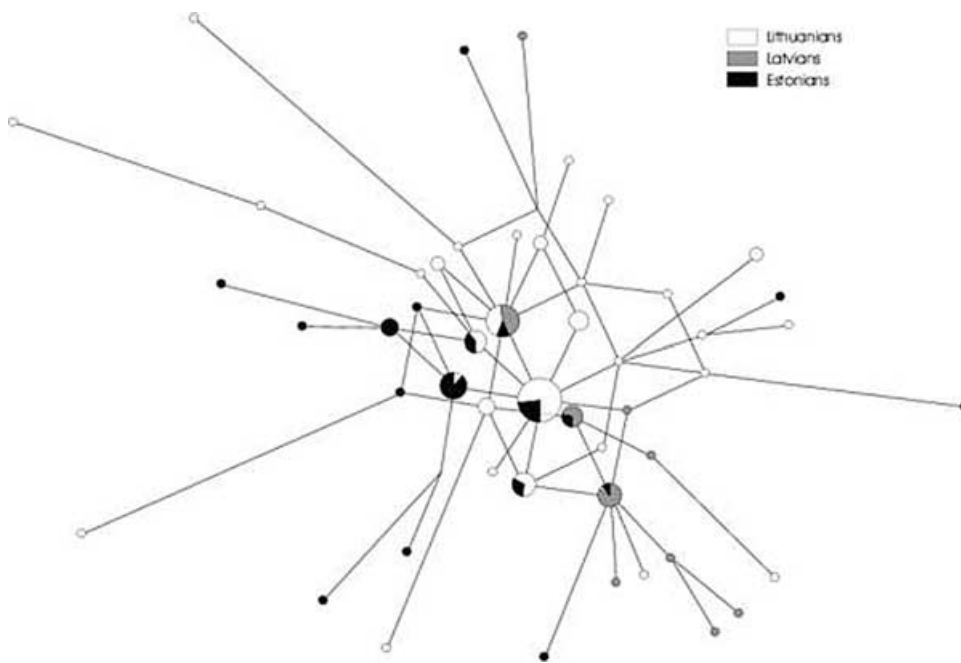

Figure 4 Median-joining network of haplogroup N3 Y chromosome STR haplotypes. Each circle represents a haplotype and the circle area is proportional to the number of chromosomes, with the shading indicating the relative frequency of that haplotype in each group.

Table 6 Demographic Inferences based on Y-chromosome STR haplotype variation

	Median (95% equal-tailed interval)			
	Mutation rate according to Kayser et al. (2000b)		Mutation rate according to Zhivotovsky et al. (2004)	
Population, Haplogroup	Population growth rate/generation $(\times 10^{-3})$	Start of population expansion years $(\times 10^3)$	Population growth rate/generation $(\times 10^{-3})$	Start of population expansion years $(\times 10^3)$
Prior probabilities*	6.9 (0.3− 36.9)	4.9 (0.1− 64.6)	6.9 (0.3− 36.9)	4.9 (0.1− 64.6)
Posterior probabilities:				
Lithuanian, HgN3	22.7 (0.8− 79.2)	0.9 (0.2− 3.3)	16.3 (4.8− 43.4)	7.6 (2.9− 24.4)
Estonian, HgN3	10.4 (0.3− 54.0)	0.9 (0.1− 4.0)	18.5 (4.7− 53.3)	8.0 (2.8− 27.1)
Lithuanian and Estonian, HgN3	30.3 (1.4− 86.2)	1.0 (0.3− 3.1)	17.9 (5.0− 49.2)	7.8 (3.1− 25.1)
Lithuanian, HgR1a	40.3 (5.2− 93.0)	1.1 (0.5− 2.6)	14.5 (4.1− 41.1)	7.8 (2.9− 24.1)
Lithuanian all Y chromosomes	78.0 (42.8− 130.4)	1.0 (0.7− 1.7)	16.4 (5.8− 39.6)	7.0 (3.0− 18.3)

∗used for calculation

time period as two independent Baltic tribes, incorporating different (now extinct) Baltic tribes during the period of consolidation of the Lithuanian state. Previous studies showed minor differences between these groups in blood group and serum markers, which might reflect differences in their original gene pools (Kučinskas, 2001). However, our results concerning mtDNA HV1 sequence and RFLP polymorphisms, and Y chromosomal biallelic and STR variation, in the Lithuanian population did not reveal any significant differences among ethnolinguistic groups of Lithuanians. Analysis of molecular variance showed that all mtDNA HV1 sequence and Y chromosome STR variation falls within the groups, indicating an extreme homogeneity of Lithuanians. Thus it is likely that, even if genetic differences between Baltic tribes existed in the past, they disappeared during the last millennium. After unification of Baltic tribes and formation of the Lithuanian state in 13th century, no strong barriers to dispersal existed within the country; however, until the end of the 19th century the admixture of people was limited because of the feudal system and serfdom. This resulted in linguistic differentiation that seems not to be reflected in the genetic composition of Lithuanians. However, the sample sizes of 30–40 individuals used in this study might be too small to exclude the existence of minor differences between the groups. Alternatively, the Baltic tribes from which modern Lithuanians originated may have been genetically homogeneous. In any event, the

molecular genetic diversity results are consistent with anthropological data, according to which anthropometric differences among regions of Lithuania disappeared in the medieval material, and the Lithuanian population is very homogeneous in the context of Eastern Europe or the whole of Europe (Česnys, 1991).

In comparisons with European populations, Lithuanians are closely related to Slavs (Russians and Poles) and Finno-Ugrians (Estonians and Finns), with respect to both mtDNA HV1 sequences and Y chromosome biallelic markers. These populations are also geographically closest among the populations analysed. Moreover, we found a significant correlation between genetic and geographic distances for both mtDNA and the Y chromosome. Since deep-rooting markers were employed in these analyses, it is hard to distinguish whether these similarities reflect common origin or later admixture of populations. On the basis of biallelic Y chromosomal markers, two main components in the Lithuanian male gene pool can be distinguished – haplogroup R1a and haplogroup N3 (TatC) Y chromosomes, comprising 45% and 37% of all Y Lithuanian chromosomes, respectively. Each of these haplogroups is defined by unique-event binary polymorphisms, and their spread is assumed to be unaffected by selection and to be the result of male migrations, influenced by local processes such as founder effects, genetic drift and gene flow.

Haplogroup R1a Y chromosomes, characterized by the recurrent SRY-1532 A>G>A transition, are found

in many European populations (Zerjal *et al.* 1999, Rosser *et al.* 2000) and are also observed at substantial frequencies in Northern India, Pakistan and Central Asia (Underhill *et al.* 1997). However, the low level of associated Y-STR diversity suggests a recent spread of this haplogroup over this wide geographic area (Zerjal *et al.* 1999). Haplogroup R1a Y chromosomes are frequent in different linguistic groups, including Indo-European (Baltic, Slavic, Hindi), Finno-Ugric, Turkic and Dravidian speakers, and hence there are no clear linguistic associations with this haplogroup. Some authors (Zerjal *et al.* 1999, Semino *et al.* 2000) suggested that, in Europe, the spread of haplogroup R1a may have been magnified by the "Kurgan culture" expansion postulated by Gimbutas (1970). This interpretation is supported by the prevalence of R1a Y chromosomes in Eastern Europe, with a decreasing westward frequency gradient (Rosser *et al.* 2000), the highest associated Y-STR diversification being detected in the Ukraine (Passarino *et al.* 2001). According to this scenario, indigenous pastoral nomadic communities, who had domesticated the horse, expanded from the forest steppes between the Dnepr and Volga rivers and surrounding areas, and brought Indo-European languages into Europe and the Indian subcontinent. In the mid-fourth millennium BP the third Kurgan wave, characterized by the Corded Ware culture, reached the present-day Lithuanian territory, where the Balts appeared via interactions with local people. Recent re-analysis of anthropological data also demonstrated probable inflows of Central-Europeans into the pre-Indo-European substratum of Lithuania (Česnys, 2001). However, Kurgan migrations are only one of several hypotheses to account for the spread of the Indo-European languages and the subject continues to be debated. The gene inflow from Central Europe to the Baltic area could have occurred during a long stage of admixture, since the transition to farming in the Baltic circum area was a slow process. It is possible that newcomers became the social elite and reduced the reproductive success of other males, and hence the Y chromosome may emphasize the real genetic contributions of the migrations from Central Europe to present-day Lithuania.

This might also explain the reduced diversity of the second major Y chromosome haplogroup – N3 – in Lithuanians. These chromosomes, defined by the Tat T>C transition, are found in Northern Eurasia in Uralic and Altaic speaking populations (Zerjal *et al.* 1997). It was suggested that they originated in Asia and were brought to Europe with Uralic speakers, prior to the arrival of Indo-European speakers. The only Indo-European speaking populations in which HgN3 chromosomes are found in frequencies comparable with those in Finno-Ugric populations, are the Balts – Lithuanians and Latvians. Zerjal *et al.* (2001) detected a genetic boundary between Baltic (Lithuanian and Latvian) and Finno-Ugric (Estonian) populations, which was determined by different patterns of Y-STR variation (primarily differences in DYS19 alleles) associated with haplogroup N3 Y chromosomes. This was interpreted as a signal of two different early migrations of the people carrying N3 Y chromosomes from Asia to Europe. This view is also supported by archaeological data (Rimantienė, 1996), odontological differences, and earlier analysis of blood serum markers (Beckman *et al.* 1998), according to which Finno-Ugrians had little influence on the formation of the Balts. Our analysis of more Y-STR markers in a much larger sample set of Lithuanians and Estonians confirms the differentiation of these populations, and the reduced Y-STR diversity associated with N3 chromosomes in Lithuanians. It is not possible to determine whether this differentiation is the result of different source populations, or if it occurred because of some period of isolation and genetic drift after an early immigration from Asia. However, since N3 and R1a chromosomes show different levels of differentiation between populations, this indicates that the populations were already differentiated before the contact of people carrying N3 and R1a chromosomes.

The higher diversity of Estonian HgN3 chromosomes suggests a different history for this population. Even though the frequency of HgR1a chromosomes is nearly the same in Estonians and Lithuanians, possibly the inflow of men carrying them did not cause a severe reduction in the autochthonous male population. If we assume that R1a chromosomes reflect an Indo-European influence, this might also explain why the forefathers of Estonians preserved a Finno-Ugric language.

Demographic analyses based on STR variation revealed a beginning of population growth between 1000 and 7000 years ago, both in Lithuanians and Estonians. The time when population growth started is the same

for HgN3 and for the other Y-chromosomes, suggesting that it was characteristic of the whole population. This wide date interval mostly reflects uncertainty about the effective mutation rate. The younger date (based on the pedigree mutation rate) would be in good agreement with archaeological and anthropological records, which show that in the first millennium AD population numbers increased due to the rise of agriculture.

Overall, the Y chromosome and mtDNA diversity show similar relationships of Lithuanians with other populations. However, we did not find reduced diversity either in the total mtDNA gene pool of Lithuanians, or within any mtDNA haplogroup. This might be due to different histories of females and males, which is also supported by the absence of a correlation between genetic distances based on mtDNA and Y chromosome variation. Perhaps immigration from Central Europe primarily involved males; also, if elite dominance played an important role, this process would have a stronger impact on Y chromosome variation. Alternatively, it might be that no significant differences in mtDNA diversity between founder populations existed in the period of formation of the Balts; therefore, we cannot distinguish different components of founders in the present day population.

In conclusion, our study demonstrates that Lithuanians are a genetically homogeneous population. Mitochondrial DNA diversity revealed that Lithuanians are close both to Slavic (Indo-European) and Finno-Ugric-speaking populations of Northern and Eastern Europe. While on the basis of biallelic Y-chromosome markers Lithuanians are closer to Finno-Ugric populations rather than other Indo-European groups, associated STR diversity indicates different histories for populations in the East Baltic area, consistent with archaeological data.

Acknowledgements

References

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R. & Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.

Bandelt, H.-J., Forster, P. & Rohl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**, 37–48.

Bandelt, H. J., Quintana-Murci, L., Salas, A. & Macaulay, V. (2002) The fingerprint of phantom mutations in mitochondrial DNA data. *Am J Hum Genet* **71**, 1150–1160.

Beckman, L., Sikström, C., Mikelsaar, A. V., Krumina, A., Ambrasienė, D., Kučinskas, V. & Beckman, G. (1998) Transferrin Variants as Markers of Migrations and Admixture between Populations in the Baltic Sea Region. *Hum Hered* **48**, 185–191.

Butrimas, A., Kazakevičius, V., Česnys, G., Balčiūnienė, I. & Jankauskas, R. (1985) Ankstyvieji virvelinės keramikos kultūros kapai Lietuvoje. *Lietuvos archeologija* **4**, 14–19. In Lithuanian.

Cavalli-Sforza, L. L. & Edwards, A. W. F.(1967) Phylogenetic analysis: models and estimation procedures. *Evolution* **32**, 550–570.

Comas, D., Calafell, F., Mateu, E., Perez-Leuzan, A., Bosch, E. & Bertranpetit, J. (1997) Mitochondrial DNA variation and the origin of the Europeans. *Hum Genet* **99**, 443–449.

Česnys, G. (1991) Anthropological roots of Lithuanians. *Science, Arts and Lithuania* **1**, 4–10.

Česnys, G. (2001) Reinventing the Bronze Age man in Lithuania: Skulls from Turlojiškė. *Acta medica Lituanica* Supplement **8**, 4–11.

Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Gimbutas, M. (1963) *The Balts*. London, New York: International Thomson Publishing.

Gimbutas, M. (1970) Proto-Indo-European culture: the Kurgan culture during the fifth, fourth and third millennia BC. In: *Indo-European and Indo-Europeans* (eds.G. Cardona, H. M. Hoenigswald & A. Senn) pp. 155–195. Philadelphia: University of Pensylvania Press.

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

Hammer, M. F. & Horai, S. (1995) Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* **56**, 951–962.

Handt, O., Meyer, S. & von Haeseler, A. (1998) Compilation of human mtDNA control region sequences. *Nucleic Acids Res* **26**, 126–129.

Harpending, H. C., Sherry, S. T., Rogers, A. R. & Stoneking, M. (1993) The genetic structure of ancient human populations. *Curr Anthropol* **34**, 483–496.

Helgason, A., Sigurethardottir, S., Gulcher, J. R., Ward, R. & Stefansson, K. (2000) mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet* **66**, 999–1016.

Helgason, A., Hickey, E., Goodacre, S., Bosnes, V., Stefansson, K., Ward, R. & Sykes, B. (2001) MtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* **68**, 723–737.

Hurles, M. E., 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, I. J., Singh, L., Pandya, A., Santos, F. R., Tyler-Smith, C. & Jobling, M. A. (1999) Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. *Am J Hum Genet* **65**, 1437–1448.

Jobling, M. A., Pandya, A. & Tyler-Smith, C. (1997) The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* **110**, 118–124.

Kasperavičiūtė, D. & Kučinskas, V. (2002) Variability of the human mitochondrial DNA control region sequences in the Lithuanian population. *J Appl Genet* **43**, 255–260.

Kayser, M., Caglià, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Schneider, P. M., Szibor, R., Teifel-Greding, J., Weichhold, G., de Knijff, P. & Roewer, L. (1997) Evaluation of Y chromosomal STRs: a multicenter study. *Int J Legal Med* **110**, 125–133, 141–149.

Kayser, M., Brauer, S., Weiss, G., Underhill, P. A., Roetwer, L., Schiefenhövel, W. & Stoneking, M. (2000a) Melanesian origin of Polynesian Y chromosomes. *Curr Biol* **10**, 1237–1246.

Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., Krüger, C., Krawczak, M., Nagy, M., Dobosz, T., Szibor, R., de Knijff, P., Stoneking, M. & Sajantila, A. (2000b) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in farther/son pairs. *Am J Hum Genet* **66**, 1580–1588.

Kittler, R., Erler, A., Brauer, S., Stoneking, M. & Kayser, M. (2003) Apparent intrachromosomal exchange on the human Y chromosome explained by population history. *Eur J Hum Genet* **11**, 304–314.

Kučinskas, V. (1994) Human mitochondrial DNA variation in Lithuania. *Anthropol Anz* **52**, 289–295.

Kučinskas, V. (2001) Population genetics of Lithuanians. *Ann Hum Biol* **28**, 1–14.

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.

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**, 232–249.

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.

Malyarchuk, B. A., Grzybowski, T., Derenko, M. V., Czarny, J., Wozniak, M. & Miscicka-Sliwka, D. (2002) Mitochondrial DNA variability in Poles and Russians. *Ann Hum Genet* **66**, 261–283.

Mathias, N., Bayés, M. & Tyler-Smith, C. (1994) Highly informative compound haplotypes for the human Y chromosome. *Hum Mol Genet* **3**, 115–123.

Miller, S. A., Dykes, D. D. & Polesky, H. F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215.

Passarino, G., Semino, O., Magri, C., Al-Zahery, N., Benuzzi, G., Quintana-Murci, L., Andellnovic, S., Bullc-Jakus, F., Liu, A., Arslan, A. & Santachiara-Benerecetti, A. S. (2001) The 49a,f haplotype 11 is a new marker of the EU19 lineage that traces migrations from northern regions of the Black Sea. *Hum Immunol* **62**, 922–932.

Redd, A. J., Clifford, S. L. & Stoneking, M. (1997) Multiplex DNA typing of short-tandem-repeat loci on the Y chromosome. *Biol Chem* **378**, 923–927.

Rimantienė, R. (1996) *Akmens amžius Lietuvoje*. Vilnius: Žiburio leidykla. In Lithuanian.

Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D. G., Brede, G., Cooper, G., Corte-Real, H. B., de Knijff, P., Decorte, R., Dubrova, Y. E., Evgrafov, O., Gilissen, A., Glisic, S., Golge, M., Hill, E. W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S. A., Krumina, A., Kucinskas, V., Lavinha, J., Livshits, L. A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T. A., Mikelsaar, A. V., Mitchell, R. J., Nafa, K., Nicholson, J., Norby, S., Pandya, A., Parik, J., Patsalis, P. C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M. J., Previdere, C., Roewer, L., Rootsi, S., Rubinsztein, D. C., Saillard, J., Santos, F. R., Stefanescu, G., Sykes, B. C., Tolun, A., Villems, R., Tyler-Smith, C. & Jobling, M. A. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* **67**, 1526–43.

Santos, F. R., Pandya, A., Tyler-Smith, C., Pena, S. D. J., Schanfield, M., Leonard, W. R., Osipova, L., Crawford,

M. H. & Mitchell, R. J. (1999) The central Siberian origin for native American Y chromosomes. *Am J Hum Genet* **64**, 619–628.

Schneider, S., Roessli, D. & Excoffier, L. (2000) Arlequin ver. 2.000: *A software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

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* **290**, 1155–1159.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphisms. *Genetics* **123**, 585–595.

The Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* **12**, 339–348.

Torroni, A., Petrozzi, M., D'Urbano, L., Sellitto, D., Zeviani, M., Carrara, F., Carducci, C., Leuzzi, V., Carelli, V., Barboni, P., De Negri, A. & Scozzari, R. (1997) Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. *Am J Hum Genet* **60**, 1107–1121.

Tyler-Smith, C. (1999) Y-chromosomal DNA markers. In: *Genomic diversity: Applications in human population genetics* (eds.S. Papiha, R. Deca & R. Chakraborty) pp. 65–73. New York: Plenum Press.

Underhill, P. A., Jin, L., Lin, A. A., Mehdi, S. Q., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. & Oefner, P. J. (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* **7**, 996–1005.

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

Wilson, I. J. & Balding, D. J. (1998) Genealogical inference from microsatellite data. *Genetics* **150**, 499–510.

Vanagas, A. (1987) Prabaltai ir baltai. Baltu arealas toponimijos duomenimis. In: *Lietuviu etnogenezė* (eds.R. Kulikauskienė, J. Jurginis, V. Mažiulis & A. Vanagas) pp. 47–62. Vilnius. In Lithuanian.

Vigilant, L., Pennington, R., Harpending, H., Kocher, T. D. & Wilson, A. C. (1989) Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci U S A* **86**, 9350–9354.

Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F. R., Schiefenhövel, W., Fretwell, N., Jobling, M. A., Harihara, S., Shimizu, K., Semjimaa, D., Sajantila, A., Salo, P., Crawford, M. H., Ginter, E. K., Evgrafov, O. V. & Tyler-Smith, C. (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* **60**, 1174–1183.

Zerjal, T., Pandya, A., Santos, F. R., Adhikari, R., Tarazona, E., Kayser, M., Evgrfov, O., Singh, L., Thangaraj, K., Destro-Bisol, G., Thomas, M. G., Qamar, R., Mehdi, S. Q., Rosser, Z. H., Hurles, M. E., Jobling, M. A. & Tyler-Smith, C. (1999) The use of Y-chromosomal DNA variation to investigate population history: Recent Male Spread in Asia and Europe. In: *Genomic diversity: Applications in human population genetics* (eds.S. Papiha, R. Deca & R. Chakraborty) pp. 91–101. New York: Plenum Press.

Zerjal, T., Beckman, L., Beckman, G., Mikelsaar, A. V., Krumina, A., Kučinskas, 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**, 1077–1087.

Zhivotovsky, L. A., Underhill, P. A., Cinnioğlu, 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

Table of mitochondrial DNA HV1 haplotype and RFLP data in Lithuanians.

Table of Y chromosome STR haplotypes in Lithuanians, Latvians and Estonians.