

Review

POPULATION GENETICS OF LATVIANS IN THE CONTEXT OF ADMIXTURE BETWEEN NORTH-EASTERN EUROPEAN ETHNIC GROUPS

Astrīda Krūmiņa¹, Liāna Pliss², Gunita Zariņa³, Agrita Puzuka¹, Agnese Zariņa¹, Baiba Lāce^{2,4}, Didzis Elferts⁵, Andrey Khrunin⁶, Svetlana Limborska⁶, Jānis Kloviņš², and Linda Gailīte (Piekuse)^{1,#}

¹ Scientific Laboratory of Molecular Genetics, Rīga Stradiņš University, Rīga, 16 Dzirciema Str., LV-1007, LATVIA

² Latvian Biomedical Research and Study Centre, 1 Rātsupītes Str., LV-1067, Rīga, LATVIA

³ Institute of Latvian History, University of Latvia, 4 Kalpaka Blvd., Rīga, LV-1050, LATVIA

⁴ Centre Hospitalier De l'Université Laval, 2705 Boulevard Laurier, Ville de Québec, QC G1V 4G2, Québec, CANADA

⁵ Faculty of Biology, University of Latvia, 1 Jelgavas Str., LV-1004, Rīga, LATVIA

⁶ Institute of Molecular Genetics, Russian Academy of Sciences, 2 Kurchatov Sq., Moscow, 123182, RUSSIA

Corresponding author, linda.gailite@rsu.lv

Contributed by Jānis Kloviņš

This article presents a review on population genetics of Latvians, which alongside Lithuanians are the two extant Baltic speaking populations. The article provides a description of genome-wide single nucleotide polymorphism (SNP) data and contains a comparative analysis of the results of studies performed on classical autosomal genetic markers, mitochondrial DNA (mtDNA) and the non-recombining part of the Y chromosome (NRY), with data on neighbouring populations. The study also covers data of recently performed ancient DNA (aDNA) studies carried out on samples from the territory of today's Latvia. The results of population genetic studies have shown a mixture of eastern and western genetic traits in present-day Latvians with only small differences between Latvian subpopulations. Studies of the Baltic "tribal gene" LW^b, as well as the gene's SERPINA1 allele PIZ have indicated the presence of a considerable Baltic admixture in the neighbouring Finno-Ugric and Slavic populations. Although mtDNA analyses have shown that Latvians genetically in general belong to the same common gene pool as most of the Europeans, the Y-chromosomal lineage composition suggests that they are most similar to Northern and Eastern European populations of Lithuanians, Estonians, and Eastern-Slavic populations, which are ethnogenetically closest to them. The analysis of aDNA from the Early and Middle Neolithic did not present any genomic evidence of gene-flow from Central European farmers or any mitochondrial or Y-chromosomal haplogroups that are typical for them in the hunter-gatherers from the territory of today's Latvia and Lithuania.

Key words: population genetics, Latvians, autosomes, mitochondrial DNA, Y chromosome.

INTRODUCTION

Contemporary people of North-Eastern European populations, including Latvians, genetically are very similar to the neighbouring populations. Principal component (PC) analysis of more than 270 000 autosomal SNPs in 3112 individuals from general populations across 16 European countries resulted in the development of genetic maps with four clusters displaying the Baltic region (Estonia, Latvia, and Lithuania) in the same cluster with Western Russia and Poland (Nelis *et al.*, 2009; see below for details). However, there are several issues concerning Latvian ethnogenesis

that remain unanswered, such as: differentiation of regional subpopulations, paternal and maternal population history, migrations affecting the region from external and internal population events, genetic admixture with neighbouring populations and identification of genetic markers that can distinguish Finno-Ugric, Slavic and Baltic influences in the Baltic Sea region.

HISTORICAL BACKGROUND

The early inhabitants of the Eastern Baltic region were reindeer hunters who arrived to this territory after the retreat of

the continental ice sheet ~12 000 years before present (YBP, all data calibrated), at the end of the Last Glacial period (Gimbutiene, 1994; Zagorska, 2012; Bērziņš and Vasks, 2013). The present-day territory of Latvia evidently represented the northernmost limit of the distribution of these hunter groups on the eastern shore of the Baltic Basin. All findings indicate that reindeer hunters reached the area of present-day Latvia from southern periglacial refuges in the course of repeated seasonal migrations. Their groups were small and mobile (Zagorska, 2012). As climatic conditions improved ~7500–10 000 YBP, the first stable settlements in the Eastern Baltic region appeared (Gimbutiene, 1994; Radiņš, 2012; Bērziņš and Vasks, 2013). The most ancient settlements in southern Lithuania are some 2000 years older than the first settlements in Latvia and Estonia (Rimantiene, 1996). The Neolithic transition in North-Eastern Europe is thought to have been slower and more gradual than in Western/Central Europe and to have involved little if any migration of early farmers from Central Europe (Jones *et al.*, 2017; Mathieson *et al.*, 2017; Mittnik *et al.*, 2017a). There is a suggestion that the Eastern Baltic was a genetic refugium for hunter-gatherer populations during the Neolithic period (Malmström *et al.*, 2009; Lazaridis *et al.*, 2014; Mittnik *et al.*, 2017a). The persistence of hunter-gatherer ancestry in the Baltic region until at least the Middle Neolithic provides a possible source for the resurgence of hunter-gatherer ancestry in Central Europe from 7000 to 5000 YBP (Haak *et al.*, 2015). From the Neolithic onwards, archaeological and historical records reveal contacts between populations of North-Eastern Europe and other groups. Finno-Ugric tribes arrived to the Baltic region from the East/South East ~6000–5000 YBP (Sedovs, 2004). These tribes are regarded to be the ancestors of modern Estonians. The next migration wave came from the South ~1000 years later. These were steppe pastoralists — South-Eastern European populations who spoke the Indo-European language — Proto-Balts who are considered as the predecessors of ancient Prussians, Yotvingians, as well as the present-day Latvians and Lithuanians (Česnys, 1991; Sedovs, 2004; Radiņš, 2012). There is a near-consensus among linguists that the Baltic and Slavic languages stem from a common root, Proto-Balto-Slavic, which separated from other Indo-European languages ~4500–7000 YBP (Bouckaert *et al.*, 2012, Haak *et al.*, 2015). The split between Baltic and Slavic branches has been dated to around 3500–2500 YBP (Gray and Atkinson, 2003; Novotna and Blažek, 2007; Bouckaert *et al.*, 2012). Proto-Balts pushed the Finno-Ugric populations up north, and in the territory of the present-day Latvia and Lithuania the Finno-Ugric and other ethnic groups were assimilated into the Indo-European-speaking main Baltic population. The major Finno-Ugric tribes that have inhabited the Baltic region are the Finns, the Estonians and the presently almost completely assimilated Livs, who previously settled around the Gulf of Riga, i.e. in present-day south-western Estonia and north-western Latvia. The expansion of the Baltic tribes northwards and eastwards marked the beginning of contacts between Baltic and Finno-Ugric populations and their languages. The contact zone was impressive — starting from



Fig. 1. The map of the territory with Baltic hydronyms (Gimbutiene, 1994).

the Baltic Sea in the West to the upper reaches of the Oka and Volga rivers in the east — see the map of Baltic hydronyms (names of rivers, lakes and other bodies of water given in Fig. 1). The eastern stem of the Balts, who did not succeed to survive till historical times, had linguistic contacts with the tribes of Mari and Mordvins (Mordvi). The influence of the Baltic languages on Finnish is evident. For instance, there are about 1.1% of Balticisms in modern Finnish (Hakulinen, 1961). The differentiation of the Baltic tribes was completed around 2000–1500 YBP. According to the available data, around 2500 YBP Finno-Ugric tribes still inhabited the northern Curonian lands up to the river Abava, the Riga shore and practically the entire territory of Latvia northwards from the Daugava River. Curonian, Semigalian, Selonian, and Latgalian tribes and languages began to merge in the southernmost areas. The fusion of these tribes gave rise to the Latvian nation (Vasks *et al.*, 1997). The Lithuanian nation was formed south of Latvia (Kučinskas, 2001). In the middle of the 1st millennium AD, Slavic tribes, representing another group of Indo-European language speakers, arrived to the Baltic territory coming along the Dnieper, Desna, and other rivers (Sedovs, 2004). The ancient Baltic territory was split into the western and eastern parts, the former comprising Old Prussian, Latvian, and Lithuanian lands, and the latter — the remaining Baltic tribes. The eastern Balts were eventually assimilated by Slavs; however, the degree of assimilation differed from region to region (Fig. 2). Thus, the contemporary population of Latvia, the focus of this review, is composed of a complex mixture of former Baltic tribes, with varying influences from Finno-Ugric and Slavic sources. Mostly due to the strategic location, since medieval times the territory of Latvia was a frequent focal point for conquests among Germany, Poland, Sweden, and Russia.

The Baltic Sea has been a considerable geographical hurdle to human migration in prehistoric times when changes in the genetic constitution due to migration were most profound. This is reflected by the differences between populations on the western (Swedish) and eastern (Finnish and Baltic) side of the Baltic Sea with respect to cultural, lin-



Fig. 2. Baltic tribes in the 13th century. Adapted from Gimbutiene (1994) and Kučinskas (2001).

guistic and genetic characteristics. The original populations on the eastern side of the Baltic Sea speak Finno-Ugric or Baltic languages, whereas those on the western side speak Germanic languages. Although the Baltic Sea has been a geographical barrier to human migration in the past, archaeological and historical evidence indicate that particularly from about 500 AD there have been lively contacts between the peoples of the Baltic Sea region.

GENETIC STUDIES OF LATVIANS

Genetic studies of Latvians were started in the first decades of the 20th century with exploration of blood groups and different anthropological features. A high frequency of blood group B (marking eastern influence) combined with high frequency of Rh-negative blood group (West European trait) was found (Prīmanis, 1937; Cauna, 1942). Contemporary molecular genetic investigations of the Latvian population were begun only ~20 years ago. Over the last decade, the ability to investigate population genetic structure has been significantly enhanced by advances in high-throughput genotyping technologies, as they allow simultaneous genotyping of hundreds of thousands of polymorphic markers.

This review deals mainly with the results obtained on unrelated healthy ethnic Latvians representing anthropologically, archaeologically, and ethno-linguistically four different regions of Latvia: Northern Courland (Lat. *Curonia*) (North-Western region, NW), Southern Courland (South-Western region, SW), Semigalia (Central region, CE) and Vidzeme/Latgalia (Lat. *Lettgallia*) (Eastern region, E) (Denisova, 1975; Denisova, 1997, Fig. 3A), who have been inhabitants of each of these regions for at least three generations.

Results of studies performed on autosomal genetic markers (genome-wide single nucleotide polymorphisms and classical autosomal genetic markers), mitochondrial DNA (mtDNA) and the non-recombining part of the Y chromosome (NRY) will be considered and compared with the data from neighbouring populations (Fig. 3B). Data on populations from some Western and Southern European countries are included mainly for comparison. The uniparentally in-

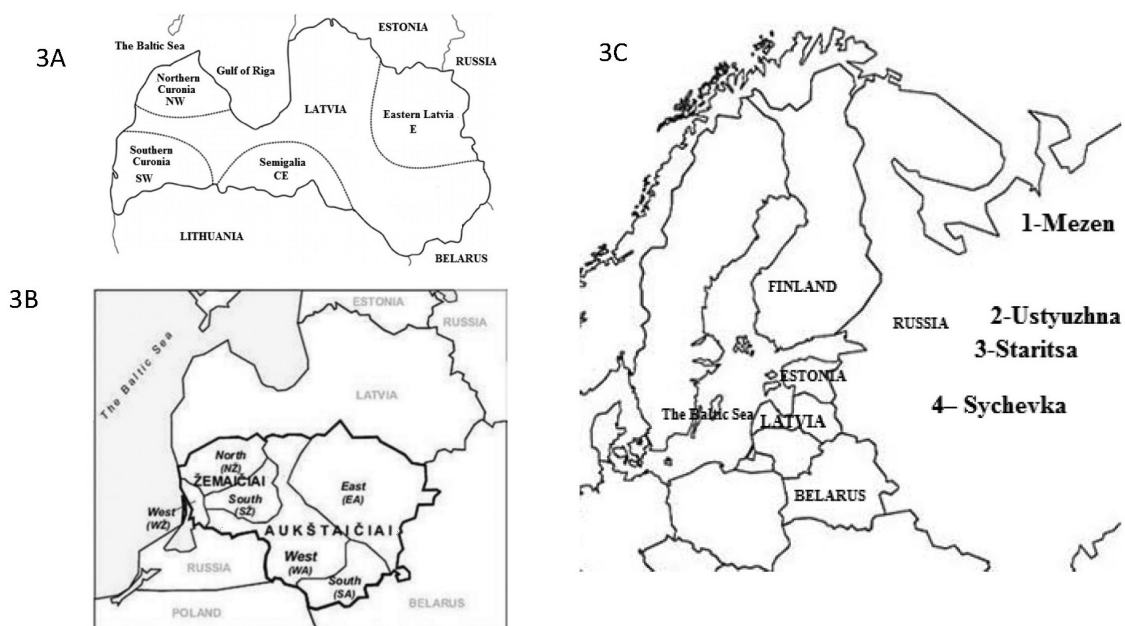


Fig. 3. Geographic location of Latvian and neighbouring subpopulations used in the reviewed population genetic studies. 3A, Latvian subpopulations: NW, North-Western region (Northern Curonia (Courland)). Ancient territory of Livonians; CE, Central region (Semigalia). Ancient territory of Semigalians; SW, South-Western region (Southern Curonia (Courland)). Ancient territory of Curonians; E, Eastern region (Vidzeme /Latgalia). Ancient territory of Latgalians; 3B, Lithuanian subpopulations (Kučinskas, 2001); 3C, Russian subpopulations: 1, Mezen district (Archangelsk Oblast); 2, Ustyuzhna district (Vologda Oblast); 3, Staritsa district (Tver Oblast); 4, Sychevka district (Smolensk Oblast).

herited mtDNA and NRY are two loci that have been extensively studied in human populations, in part because they represent maternal and paternal population histories and in part because they do not undergo recombination and, therefore, lineages can be more easily traced back to a single common ancestor. There are differences in the action of genetic drift and natural selection on mtDNA and NRY loci as compared to autosomal loci. Thus, a combination of mtDNA and NRY loci with autosomal data is necessary for a thorough understanding of population history.

Results of studies on ancient DNA samples are also included because analysis of ancient DNA can provide information on past population histories that is not retrievable from contemporary individuals.

In this review, besides a description of genome-wide SNP data, we will also focus on several autosomal genes. These genes are chosen because of known regional frequency differences in North-Eastern Europe. They give information not only on the history of certain mutations, but also on population migration routes in Eastern Baltic, possible selective advantages of these mutations and their contribution to diseases. For two autosomal diseases (cystic fibrosis and phenylketonuria), phenotype (not allelic) frequencies have been compared. Frequencies of the mentioned diseases are given for the general (not ethnic) population.

AUTOSOMAL GENETIC MARKERS

Genome-wide single nucleotide polymorphisms (SNPs).

The ability to investigate population genetic structure has been significantly improved over the last decade by advances in high-throughput genotyping technologies. Genome-wide analyses in human populations allow to unveil the demographic history underlying the genetic patterns of diversity observable today across the globe. In Europe the strongest genetic differentiation has been found between the north-west and south-east parts of the continent (Li *et al.*, 2008). Populations from North-Eastern Europe, incl. Latvians, are less represented in these studies.

As already mentioned at the beginning of the introductory part, principal component (PC) analysis (PCA) of more than 270 000 autosomal SNPs genotyped with the Illumina Infinium platform in 3112 individuals from general (but not always ethnic) populations across 16 European countries, including 95 individuals from Latvia, resulted in a boomerang-like structured genetic map of Europe with four clusters displaying individuals from the Baltic region (Latvia, Lithuania, and Estonia) in the same cluster with Polish and western Russian samples (Fig. 4; Nelis *et al.*, 2009).

Khrunin *et al.* (2013) examined the European population structure with 166 000 autosomal SNPs and obtained another genetic map of Europe (some populations from North-Eastern regions of the European part of Russia, Komi, Veps, Northern Russians were also included in the study). In the space of the plotted first two principal components,

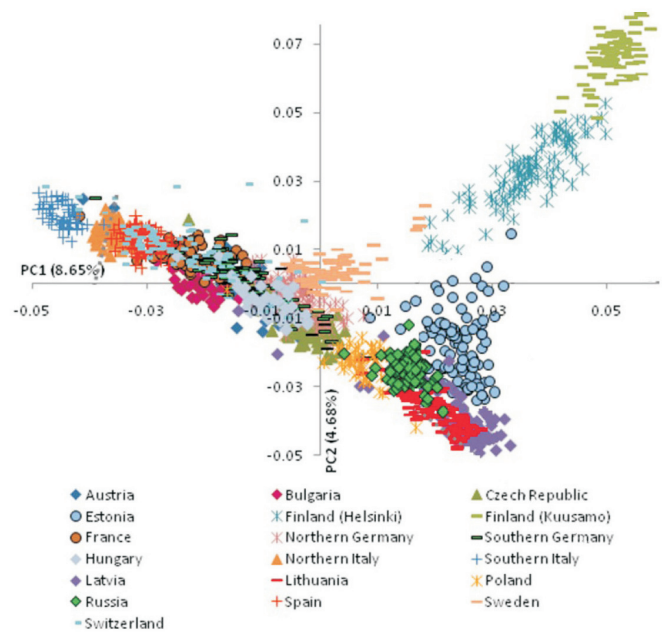


Fig. 4. The European genetic structure (based on 273 464 SNPs) as revealed by Principal Component Analysis (Nelis *et al.*, 2009).

Latvians clustered in close proximity to populations from Central and Eastern Europe, demonstrating the highest similarity with autosomal SNP profiles of Russians from the central part of European Russia and Estonians (Lithuanians were not included in the study). The relationships found were similar with the results of ADMIXTURE ancestry analysis, revealing a particular specificity of the Latvian gene pool, which exhibited a lower proportion of the “Finnic” ancestry component than in Estonians. Additionally, Latvians had the highest proportion of non-specified yellow component (there were no populations in the data set that had only this component in their genomes), which was omnipresent throughout the studied European populations (Khrunin *et al.*, 2013).

Similar results of the genetic analysis of Balto-Slavic speaking populations were obtained by Kushniarevich *et al.* (2015), who demonstrated the proximity of autosomal genome-wide SNP profiles of Latvians (six individuals analysed) and Lithuanians to East Slavs and Estonians. The results showed that Latvians and Lithuanians shared higher number of segments that were identical by descent with the combined group of east-west Slavs than in other non-Slavic people did (Kushniarevich *et al.*, 2015).

Results on high-coverage whole-genome sequence studies performed on three present-day Latvians are included in the dataset of Estonian Biocentre Human Genome Diversity Panel (Pagani *et al.*, 2016). The dataset was analysed for numerous parameters, including shared ancestry, archaic admixture, and changes in effective population size through time. These analyses have shown that Latvians are located in the North European genetic cluster with highest similarity to Lithuanians.

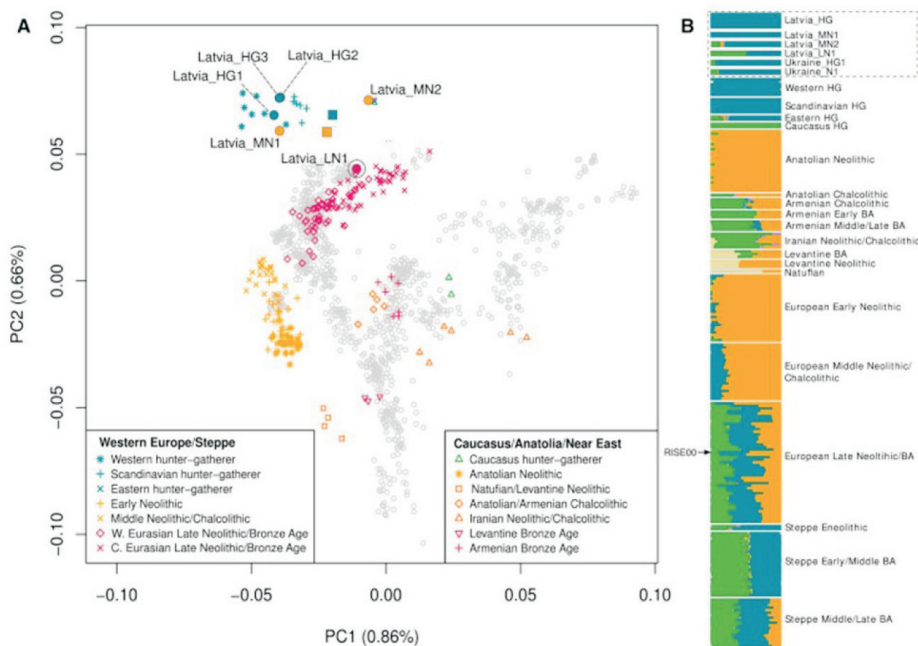


Fig. 5. Principal component analysis (PCA) and ADMIXTURE analysis of ancient Latvian DNA samples (adapted from Jones *et al.*, 2017). 5A, ancient DNA data presented in the study as well as published ancient DNA data are projected onto the first two principal components defined by selected modern Eurasian data; 5B, ADMIXTURE ancestry components for ancient DNA samples showing that the Latvian Neolithic samples do not have the yellow component that dominates in Anatolian and early European farmers. The Latvian and Ukrainian samples presented in the study are displayed in a gray box and at twice the height of the other ancient samples for ease of visualization. The arrow shows an Estonian Bronze Age sample (RISE00) (Allentoff *et al.*, 2015) that has a yellow component. MN, Middle Neolithic; LN, Late Neolithic; N, Neolithic; HG, hunter-gatherer; BA, Bronze Age; W, Western; C, Central.

As it was already mentioned in the Introduction, information on past population histories can be provided by genomes obtained from ancient biological remains. Recently performed genome-wide ancient DNA (aDNA) studies have revealed a complex population history of modern Europeans involving at least three major prehistoric migrations (Lazaridis *et al.*, 2014; Allentoft *et al.*, 2015; Haak *et al.*, 2015; Saag *et al.*, 2017). While these events are relatively well described in Central and Southern Europe, their genetic impact on North-Eastern Europe is not so extensively studied.

Data of recently performed genome-wide analyses of 30 Latvian ancient DNA samples are presented in Table 1. PCA comparison of 15 Latvian Mesolithic and Early Neolithic samples with data of ancient Western European hunter-gatherers (WHG) and Eastern European hunter-gatherers (EHG) as well with the genomic data of modern Europeans showed clustering of ancient Latvian samples between WHG and EHG (with closer affinity to WHG than to EHG) as well as their location close to modern Northern and Eastern European populations (Jones *et al.*, 2017; Mathieson *et al.*, 2017; Fig. 5). These data allow to suggest some genetic continuity in this region since the Mesolithic period. Similar results were obtained also for ancient Lithuanian samples (Mathieson *et al.*, 2017; Mittnik *et al.*, 2017a). Remarkable changes in the genetic data are observed in the samples from the period between the Early Neolithic and the Middle Neolithic, when clear influence from exogenous Northern Eurasian populations is observed (Jones *et al.*, 2017; Mathieson *et al.*, 2017), and also, Late Neolithic and Bronze Age samples show impact from the Pontic Steppe populations (steppe pastoralists). There are no signs of Anatolian farmer-related genetic input on populations of Latvia and Lithuania at least until the end of the Bronze Age (Jones *et al.*, 2017; Mathieson *et al.*, 2017; Mittnik *et al.*, 2017a). Remarkably, the Steppe-related genetic impact coincides with the proposed emergence of

Indo-European languages in the Baltic region (Bouckaert *et al.*, 2012; Haak *et al.*, 2015).

Genome-wide analysis of polymorphisms presents a general conception about relationships and evolutionary trajectories of population formation. In addition, population analysis of local genomic regions may provide information about historical traces of particular lineages within populations.

Gene *CFTR*, allele $\Delta F508$. Gene *CFTR* (OMIM *602421) encodes protein CFTR — cystic fibrosis transmembrane conductance regulator, which acts as a chloride ion channel. Mutations in the gene cause cystic fibrosis (CF; OMIM #219700), one of the most common severe autosomal recessive disorders in populations of Northern European ancestry in most countries, affecting 1 in 2000 to 4000 individuals (Table 2). Gene mutations affect the salt and water balance of the mucus layer lining epithelial surfaces mostly in the lungs and pancreas. The major cystic fibrosis mutation is $\Delta F508$ (delta-F508) deletion of three base pairs at codon 508, which results in a loss of the amino acid phenylalanine in the protein. The incidence of $\Delta F508$ carriers among healthy Latvians is 1 : 42, which corresponds to the frequency of $\Delta F508$ mutation carriers in most North European populations (Krumina *et al.*, 2001b). This mutation was found in 61.3% of chromosomes in the overall CF patient population in Latvia. The incidence of CF as well as of $\Delta F508$ varies markedly among different populations. There is a north-west/south-east gradient in the incidence of CF throughout Europe, with the highest incidence found in Ireland and England (1/1353 and 1/2381; Farrell, 2008), and the lowest in Finland (1/25 000; Farrell, 2008). To trace the history of $\Delta F508$ in Latvia the association of $\Delta F508$ with five polymorphic DNA markers was studied. The haplotypes observed on $\Delta F508$ CF chromosomes were the same as in other populations. Distribution of haplotypes found in the Latvian population, together with frequencies of $\Delta F508$ mutation, is intermediate between Northern- and South-

Table 1

RESULTS OF GENOME ANALYSIS FROM ANCIENT LATVIAN DNA SAMPLES

Sample's code	Cemetery/burial ⁶	Sample's age (cal YBP)*	Cultural attribution	Mean genome coverage	Gene's <i>MCM6</i> ancestral allele	mtDNA haplogroup	NRY haplogroup
Zvej30 ²	Zvejnieki 305	8240 ± 70	Mesolithic hunter-gatherers	2.43	nda	U5a2c	R1b1a1a
Latvia HG1 ¹ , Zvej32 ²	Zvejnieki 313	7525 ± 60	Mesolithic hunter-gatherers	1.38 ¹ , 3.51 ²	+ + ⁴	U5a1c ^{1,2}	Female
Zvej10 ²	Zvejnieki 67	7140 ± 40	Mesolithic hunter-gatherers	3.76	nda	U5a2c	R1b1a1a
Latvia HG2 ¹ , Zvej25 ²	Zvejnieki 93	6840 ± 55	Mesolithic hunter-gatherers	4.37 ¹ , 0.54 ²	+ + ⁴	U2e1 ^{1,2}	R1b1b ¹ , R1b1a1a ²
Zvej20 ²	Zvejnieki 86	6815 ± 40	Mesolithic hunter-gatherers	3.78	nda	U5b1d1	R1b1a1a
Zvej12 ²	Zvejnieki 128	6530 ± 35	Mesolithic hunter-gatherers	3.48	nda	U5a2d	R1b1a1a
Zvej3 ²	Zvejnieki 52	Not dated	Mesolithic hunter-gatherers	4.42	nda	U5b2a1a	Q1a2
Zvej4 ²	Zvejnieki 108	Not dated	Mesolithic hunter-gatherers	3.90	nda	U5a2c3	I2a2a1
Zvej5 ²	Zvejnieki 117	Not dated	Mesolithic hunter-gatherers	4.17	nda	U5a1d	Female
Zvej7 ²	Zvejnieki 98	Not dated	Mesolithic hunter-gatherers	1.03	nda	U5a1c	I2a2a1
Zvej16 ²	Zvejnieki 224	6335 ± 35	Mesolithic/Neolithic hunter-gatherers	3.82	nda	U4b1a2	I2a1
Zvej21 ²	Zvejnieki 197	6180 ± 30	Mesolithic/Neolithic hunter-gatherers	3.22	nda	U4a1	I2a1
Latvia HG3 ¹ , Zvej27 ²	Zvejnieki 121	6145 ± 55	Mesolithic/Neolithic hunter-gatherers	0.97 ¹ , 2.88 ²	+ - ⁴	U5a2d	R1b1b ¹ , R1b1a1a ²
Zvej22 ²	Zvejnieki 173	5900 ± 30	Early Neolithic hunter-gatherers	3.52	nda	U5a1c	I2a2a1b
Zvej15 ²	Zvejnieki 226	5370 ± 30	Early Neolithic hunter-gatherers	3.60	nda	U5a1d2	Female
Zvej14 ²	Zvejnieki 261	5335 ± 30	Middle Neolithic hunter-gatherers	3.30	nda	U4a1	R1b1a1a
Latvia MN1 ¹ , Zvej26 ²	Zvejnieki 124	5280 ± 55	Middle Neolithic hunter-gatherers	0.18 ¹ , 1.77 ²	nda	U4a1	nda ¹ , R1b1a1a ²
Zvej13 ²	Zvejnieki 278	5 210 ± 30	Middle Neolithic hunter-gatherers	3.41	nda	U4a1	Female
Latvia MN2 ¹	Zvejnieki 221	5180 ± 65	Middle Neolithic. Comb Ware culture	1.50	+ + ⁴	U4	Female
Zvej24 ²	Zvejnieki 207	5180 ± 30	Middle Neolithic hunter-gatherers	2.53	nda	U4a1	Female
Latvia LN ¹ , Zvej28 ²	Zvejnieki 137	4280 ± 60	Late Neolithic. Corded Ware culture	0.26 ¹ , 0.32 ²	nda	U5a1 ¹ , U5a1b ²	Female
Kivutkalns42 ³	Kivutkalns 42	2573 ± 30	Late Bronze Age	1.10	+ - ⁵	H1b1	Female
Kivutkalns25 ³	Kivutkalns 25	2545 ± 30	Late Bronze Age	1.57	- - ⁵	H28a	R1a1a1b
Kivutkalns153 ⁷	Kivutkalns 153	2542 ± 29	Late Bronze Age	0.33	+ +	U5a1a1	R1b1a2
Kivutkalns207 ³	Kivutkalns 207	2511 ± 30	Late Bronze Age	7.21	+ - ⁵	H1b2	Female
Kivutkalns209 ³	Kivutkalns 209	2497 ± 30	Late Bronze Age	2.24	+ - ⁵	J1b1a1	R1a1a
Kivutkalns215 ³	Kivutkalns 215	2462 ± 27	Late Bronze Age	2.74	+ - ⁵	H1c	Female
Kivutkalns222 ³	Kivutkalns 222	2423 ± 26	Late Bronze Age	1.28	+ - ⁵	U5a1c1	R1a1
Kivutkalns19 ³	Kivutkalns 19	2403 ± 24	Late Bronze Age	5.76	+ + ⁵	H10a	R1a1a1b
Kivutkalns194 ³	Kivutkalns 194	2298 ± 28	Late Bronze Age	0.15	- - ⁵	T1a1b	R1a1a

Notes: *cal YBP, calibrated years before present; nda, no data available; ¹Jones *et al.*, 2017; ²Mathieson *et al.*, 2017; ³Mittnik *et al.*, 2017a; Jones *et al.*, 2017; ⁴ancestral allele 22018G (rs182549) (Mittnik *et al.*, 2017b); ⁵ancestral allele 13910C (rs4988235); ⁶Zarina's data; ⁷Mittnik *et al.*, 2017b.

Table 2

INCIDENCE OF CYSTIC FIBROSIS AND PHENYLKETONURIA IN EUROPEAN POPULATIONS

Populations	Phenylketonuria (PKU)	R408W (rs5030858) frequency (%) of PKU alleles	Cystic fibrosis (CF)	$\Delta F508$ (rs113993960) frequency (%) of CF alleles
Latvians	1/ 8000 ¹	73.0	1/ 3300 ⁴	61.3
Lithuanians	1/ 10 000 ³	73.4	1/ 2500 ⁵	34.0
Russians	1/ 10 000 ³	56.5	1/ 4900 ⁷	54.5
Belarusians	1/ 5600 ³	68.2	1/ 8000 ⁸	61.2
Ukrainians	1/ 8300 ³	57.0	1/3300 ⁵	65.2
Poles	1/ 8000 ³	62.2	1/ 5000 ⁵	57.1
Czechs	1/ 7000 ²	61.1	1/2833 ⁵	70.0
Danes	1/ 12 000 ²	18.2	1/4700 ⁵	87.5
Norwegians	1/ 14 500 ²	13.8	1/4500 ⁶	60.2
Irish	1/ 4500 ³	42.5	1/1353 ⁵	70.4
Swedes	1/ 20 000 ³	21.9	1/5600 ⁵	66.6
English people	1/ 12 000 ³	11.1	1/2381 ⁶	75.3
Germans	1/ 8000 ³	25.7	1/ 3 300 ⁵	71.8
Italians	1/ 17 000 ²	1.4	1/4238 ⁵	50.9
French	1/ 13 500 ²	14.1	1/4700 ⁵	67.7
Estonians	1/ 8090 ³	84.0	1/ 4500 ⁵	51.7
Finns	1/ 150 000 ³	68.0	1/ 25 000 ⁵	46.2

¹ Purina *et al.*, 1995; ² Williams *et al.*, 2008; ³ Tighe *et al.*, 2003; ⁴ Krumina *et al.*, 2000; ⁵ Farrell, 2008; ⁶ World Health Organization; ⁷ Stepanova *et al.*, 2016; ⁸ Mosse and Mosse, 2005

ern-European populations. These results are consistent with the hypothesis of a single origin and subsequent diffusion of this major CF mutation from North-Western Europe. The estimate of the number of generations elapsed since the mutation $\Delta F508$ was first introduced in Latvians is approximately 101, or 2020 years (Krumina *et al.*, 2001b).

Gene PAH, allele R408W. Gene *PAH* (OMIM *612349) encodes enzyme phenylalanine hydroxylase (PAH), which catalyses hydroxylation of phenylalanine to tyrosine. Mutations in the gene cause phenylketonuria (PKU; OMIM #261600), one of the most common inborn errors of metabolism in people of European ancestry, with a frequency of 1/8000 in Latvia (Purina *et al.*, 1995; Proņina and Lugovska, 2011; Table 2). It is an autosomal recessive disease characterised by a progressive neurodegenerative course. Over 900 mutations causing PKU have been reported in the *PAH* gene. There are marked differences in the spectrum of *PAH* mutations between European countries. The most prevalent mutation R408W is found at relative allele frequencies as high as 84% in Europe. In Europe, the R408W mutation is observed on chromosomes of two major haplotype backgrounds: R408W-2.3 and R408W-1.8 (Tighe *et al.*, 2003). R408W-2.3 exhibits a gradient of allele frequency increasing eastwards, reaching its maximum in the Balto-Slavic region. R408W-1.8 exhibits an east-to-west cline in North-Western Europe, peaking in the most westerly province of Ireland. PKU in Latvia is genetically relatively homogeneous. The most common mutation R408W

Table 3

FREQUENCIES OF GENE *SERPINA1* ALLELES PIZ AND PIS IN EUROPEAN POPULATIONS

Populations	n	<i>SERPINA1</i> gene allele PIZ (rs28929474)	<i>SERPINA1</i> gene allele PIS (rs17580)
Latvians ^{1,2}			
NW	68	0.04	0.02
CE	68	0.02	0.04
SW	69	0.04	0.05
E	65	0.04	0.03
Total	270	0.03	0.03
Lithuanians ²			
Žemaitija	418	0.01	0.02
Aukštaitija	496	0.01	0.01
Total	914	0.01	0.02
Russians (NW) ³			
Staritsa, Tver Oblast	99	0.02	0.01
Sychevka, Smolensk Oblast	111	0.01	0.000
Ustyuzhna, Vologda Oblast	98	0.03	0.02
Mezen, Archangelsk Oblast	66	0.04	0.02
Total (NW)	374	0.02	0.01
Russians (Central) ⁴	2787	0.01	0.01
Belarusians ³	98	0.02	0.02
Poles ⁴	9539	0.01	0.02
Swedes (Central, S) ²	604	0.03	0.01
Swedes (Gotland) ²	574	0.03	0.03
Danes ⁴	1096	0.03	0.03
Norwegians ⁴	4492	0.02	0.03
Irish ⁵	1100	0.02	0.05
English people ⁴	4775	0.01	0.03
Germans ⁴	5886	0.01	0.02
French ⁴	8753	0.01	0.08
Hungarians ⁴	1036	0.01	0.03
Dutch ⁴	2539	0.01	0.02
Estonians ²	1850	0.03	0.01
Finns ⁴	2112	0.01	0.01

¹ Maliseva *et al.*, 2002; ² Beckman *et al.*, 1999; ³ Gailite (Piekuse), unpublished data; ⁴ Blanco *et al.*, 2006; ⁵ Carroll *et al.*, 2011

accounts for 73% of Latvian PKU chromosomes. The same prevalence of this mutation is found in the Lithuanian population and is even higher (84%) in the Estonian population. R408W-2.3 is the main haplotype for mutation R408W not only in Latvians and Lithuanians, but also in Estonian, Polish, and German populations. The frequency of R408-2.3 suggests a Balto-Slavic origin of this R408W mutation or its association with a pre-Indo-European founder population (Tighe *et al.*, 2003; Proņina and Lugovska, 2011).

Gene *SERPINA1*, alleles PIZ and PIS. Gene *SERPINA1* (OMIM *107400) encodes protein alpha-1-antitrypsin (AAT), a major plasma serine protease inhibitor. Alpha-1-antitrypsin deficiency (OMIM #613490) is one of the most common autosomal recessive inborn errors of metabolism in people of European ancestry, which is caused by muta-

tions in the gene *SERPINA1*. AAT deficiency often is associated with an early onset of obstructive lung disease and emphysema in adulthood, and liver disease in children. The most frequent AAT deficiency alleles (mutations) are PIZ and PIS. The most common mutation presenting with clinical evidence is the Z mutation (E342K), caused by a single amino acid replacement of glutamic acid at position 342 of the polypeptide, while S mutation is associated with a milder plasma deficiency of AAT protein. Throughout Europe the frequency of Z and S mutations in healthy individuals varies widely between countries. ~3–4% of North-Europeans carry the Z allele whereas 6% carry S allele. Frequency of the PIZ allele in Latvia detected by isoelectric focusing in 488 individuals was 0.040 (Beckman *et al.*, 1999), but by more precise DNA analysis in 270 individuals gave an estimate of 0.034 (Mališeva *et al.*, 2002; Table 3), which are the highest population frequencies reported so far in Europe. The calculated frequency of PIZ carriers is 1 : 30; approximately 3% of the Latvian population are carriers of the mutation PIZ. Frequencies of the PIS allele in Latvia were estimated to be 0.026 (isoelectric focusing, Beckman *et al.*, 1999) and 0.032 (DNA analysis, Mališeva *et al.*, 2002; Table 3), which are higher than the average allele PIS frequency in Central Europe, but significantly lower than in South European countries. Noteworthy is the fact that the highest frequency of PIZ allele is reported in the North-Western part of Latvia (Northern Curonian region). High frequencies are also found on the Swedish mainland and on the Island of Gotland. Relatively high frequencies are observed in Estonians, as well as in the North-Western regions of Archangelsk and Vologda in Russia (Gailite (Piekuse), unpublished data), whereas in Finns and people of Central European part of Russia the frequencies are considerably lower. There is a remarkable and highly significant difference between Latvians and Lithuanians with respect to PIZ allele frequency. In this context it is of interest that archaeological data have shown that from the 7th century and for several hundreds of years there were several settlements of people from Swedish mainland as well as from the Island of Gotland in the Grobiņa area on the Curonian coast of the Baltic Sea (Radiņš, 2012). The PIZ variant in Latvia and Sweden shared a unique genotype that was not present in the control population (Lace *et al.*, 2008). This single genotype confirmed the hypothesis of a recent and single origin of the PIZ mutation. Analysis of non-recombinant single nucleotide polymorphisms (SNPs) was used to estimate that the PIZ mutation age was 2902 years in Latvia and 2362 years in Sweden. The authors speculate that the possible route of this allele was from Sweden (southern Scandinavia) to other countries, including the Baltic countries, where it spread through the waterways into the continent by migration patterns such as the Viking colonisation of North-Western Europe between 800 and 1200 AD (Hutchison, 1998). In any case, the allele PIZ is a good marker for Southern Sweden and Latvian genetic influence. The S mutation is older than the Z mutation and is postulated to have arisen in the north of the Iberian Peninsula and subsequently spread throughout Europe during the mass migration (Seixas *et al.*, 2001).

Gene *HFE*, alleles C282Y and H63D. Gene *HFE* (OMIM *613609) codes for HFE protein that participates in the regulation of iron absorption. Mutations in this gene may lead to hereditary haemochromatosis (HH s. *HFE*-associated hereditary haemochromatosis, OMIM #235200), an inborn error of iron metabolism. HH belongs to the most common autosomal recessive disorders in people of European ancestry, characterised by iron overload and failure of different organs. Most HH patients are homozygotes for the missense mutation C282Y that leads to the substitution of tyrosine for cysteine at amino acid 282. The role of another missense mutation, H63D, is still unclear (compound heterozygotes with C282Y have haemochromatosis). The average frequencies of mutation C282Y and H63D carriers among healthy Latvians are 3.5% and 12.1%, respectively, which is similar to the average frequency in Lithuanians and Estonians, but lower than in most populations of Northern Europe (Krumina *et al.*, 2001a; Kucinskas *et al.*, 2012; Table 3). The frequency of C282Y mutation is significantly higher among the inhabitants of Žemaitija at the Baltic Sea area (5.9%) in comparison to the regions of the continental part of Lithuania (Kučinskas *et al.*, 2012). The highest frequency is observed in Ireland (11.4%) and in the eastern part of England (8.2%), and a slightly lower frequency in Norway, Sweden, Denmark, and Iceland. The lowest reported frequencies (0–1.75%) of the C282Y mutation were observed in Greece, Romania, and Bulgaria (Merryweather-Clarke *et al.*, 1997; Ivanova *et al.*, 1999; Voicu *et al.*, 2009). These data support the hypothesis of gene flow from north to south of Europe. Most scientists agree that mutation C282Y appeared ~60–70 generations ago (Ajioka *et al.*, 1997) and originated from the population of Southern Scandinavia or Northern Germany (Distante *et al.*, 2004). It has also been suggested that Viking migrations were largely responsible for the distribution of this mutation. The H63D mutation is more common and equally distributed among European nations. The highest frequency of this mutation has been reported in the Basque population (30.4%) and genetic data suggest that this mutation could have originated in early Western European populations (de Juan *et al.*, 2001).

Gene *CCR5*, allele CCR5-Δ 32. Gene *CCR5* (OMIM *601373) encodes cell membrane protein C-C chemokine receptor 5 (CCR5), which is the principal entry co-receptor for macrophage-tropic human immunodeficiency virus type 1 (HIV-1) strains. Mutation CCR5-Δ32 (CCR5-delta32) is a 32 bp deletion in this gene and results in a frame shift in the coding sequence that produces a non-functional protein, and consequently it is not expressed in the cell membrane. Individuals homozygous for the mutation show high degree of resistance to HIV infection. Heterozygous carriers have reduced susceptibility to infection and reveal a slowed progression towards AIDS. The results of screening of more than 2500 individuals from 18 European populations suggested a unique and relatively recent origin of this allele, which probably arose in North-Eastern European populations (Libert *et al.*, 1998; Klitz *et al.*, 2001). The age of the Δ32 allele has been estimated between 700 and 3500 years based on linkage disequilibrium data (Stephens *et al.*, 1998;

Table 4

FREQUENCIES OF GENE *HFE* ALLELES C282Y AND H63D IN EUROPEAN POPULATIONS

Populations	n	<i>HFE</i> gene allele C282Y (rs1800562)	<i>HFE</i> gene allele H63D (rs1799945)
Latvians ¹			
NW	78	0.03	0.10
CE	71	0.03	0.11
SW	71	0.04	0.15
E	75	0.05	0.12
Total	295	0.04	0.12
Lithuanians ²			
Žemaitija	102	0.06	nda
East Aukštaitija	378	0.02	nda
West Aukštaitijai	148	0.02	nda
South Aukštaitijai	383	0.02	nda
Total	1011	0.03	0.16
Russians (Central Eur. part) ²	150	0.04	0.13
Belarusians ³	736	0.04	0.16
Ukrainians ⁴	3300	0.03	0.17
Poles ²	871	0.03	0.16
Czechs ²	481	0.03	0.15
Danes ⁵	2501	0.06	0.13
Norwegians ¹	94	0.06	0.14
Irish ²	1002	0.11	0.15
Swedes (Central) ²	250	0.06	0.11
English people ⁶	91	0.08	0.13
Germans ¹	115	0.04	0.15
Italians ¹	141	0.01	0.13
French ¹	398	0.03	0.17
Estonians ¹	519	0.04	0.14
Finns ¹	173	0.05	0.12

nda, no data available; ¹ data from Krumina *et al.*, 2001a; ² data from Kucinkas *et al.*, 2012; ³ Sivitskaya *et al.*, 2013; ⁴ Pampukha *et al.*, 2003; ⁵ Milman *et al.*, 2004; ⁶ 1000 Genomes Browser

Slatkin, 2001), and ancient DNA evidence suggests that the allele is at least 2900 years old (Hummel *et al.*, 2005). The highest frequencies were registered in Mordvins, Estonians, Finns, Belarusians, Latvians, as well as in North-European Russians, the last populations being supposed to have a strong Finno-Ugric background, maximum in areas surrounding the Baltic and White Seas (the area of highest frequency lies between the Baltic and the White Seas; Table 5). High frequency is registered also in central Russia near Novosibirsk (Balanovsky *et al.*, 2005). If $\Delta 32$ were neutral, population genetics theory, given allele's frequency, predicts that it would have to be much older. An alternative explanation is that $\Delta 32$ mutation occurred recently and afterwards increased rapidly in frequency because of a strong selective advantage (Libert *et al.*, 1998; Stephens *et al.*, 1998). Bubonic plague was initially proposed as the selective agent, but subsequent analysis suggested that a disease like smallpox is a more plausible candidate (Galvani and Slatkin, 2003; de Silva and Stumpf, 2004; Stumpf and

Table 5

FREQUENCIES OF GENE *CCR5* ALLELE *CCR5*- $\Delta 32$ IN EUROPEAN POPULATIONS

Populations	n	<i>CCR5</i> gene allele $\Delta 32$ (rs333)
Latvians ¹		
NW	91	0.12
CE	70	0.19
SW	51	0.15
E	68	0.17
Total	156	0.15
Lithuanians ²	283	0.12
Russians (NW) ¹		
Staritsa, Tver Oblast	97	0.11
Sychevka, Smolensk Oblast	86	0.14
Ustyuzhna, Vologda Oblast	95	0.10
Mezen, Archangelsk Oblast	123	0.10
Total (NW)	401	0.11
Russians (Moscow) ²	50	0.14
Russians (Ryazan) ²	78	0.12
Belarusians ²	80	0.16
Ukrainians ²	83	0.11
Poles ³	1005	0.12
Czechs ³	547	0.11
Danes ²	100	0.12
Norwegians ³	183	0.12
Irish ³	191	0.08
Swedes ²	204	0.14
English people ³	705	0.12
Germans ³	307	0.11
Italians ²	91	0.06
French ²	276	0.09
Mordvins ²	86	0.16
Estonians ³	662	0.14
Finns ²	98	0.16

¹ Gailite (Piekuse) unpublished data; ² Limborska *et al.*, 2002; ³ Faure and Royer-Carenzi, 2008

Wilkinson- Herbots, 2004; Galvani and Novembre, 2005). Analysis performed on *CCR5* polymorphisms in Russian, Ukrainian, and Moldavian populations showed a strong positive correlation between allele frequency and latitude, and a set of temperature factors (Limborska *et al.*, 2002). The authors proposed that the existence of correlations between the cline of $\Delta 32$ frequencies and climatic-geographic parameters provided evidence for a possible effect of either natural environmental factors (geographical gradient in selection intensity) or large-scale population movements on distribution of this allele. Other authors (Balanovsky *et al.*, 2005) proposed a rise and initial spread of the mutation among Uralic-speaking populations, in particular in its Finnic branch, a frequency increase in North-Eastern Europe as a result of selection and/or genetic drift, and secondary spread (with selection continued) due to gene flow and the migrations of North Europeans across the globe.

Gene *ICAM4*, allele *LW^b*. *ICAM4* (*LW*) gene (OMIM *614088) encodes single-pass membrane glycoprotein called intercellular adhesion molecule 4. Landsteiner-Wiener (*LW*) blood group antigens (OMIM #111250) reside on this protein. There are three antigens assigned to the *LW* system: *LW^a*, *LW^b*, and *LW^{ab}* (Byrne and Byrne, 2004; Srivastava *et al.*, 2014). *LW^a* and *LW^b* antigens are antithetical, resulting from a single nucleotide substitution in the *ICAM4* gene (299A>G), which causes an amino acid change in *LW* glycoprotein (Q100R).

LW blood group polymorphism has been widely tested by antigen detection (Sistonen *et al.*, 1999). Population genetic studies on the *LW* blood group by standard serological methods resulted in a conclusion that the *LW^b* antigen was found in Europeans only, presenting the highest frequency in the Baltic-speaking Latvian and Lithuanian populations, leading to the conclusion that this polymorphism can be considered as a “Baltic tribe marker” (Kučinskas *et al.*, 1998; Sistonen *et al.*, 1999; Kučinskas, 2001). Recent advances in DNA analysis allowed to trace *LW^b* allele frequency distribution for several populations, not previously included in studies by serological methods (the 1000 Genomes Project). The *LW^b* allele was not found or was found at very low frequencies in most African, Native American, Asian, and West-European populations. Frequencies of *LW^b* allele in different European populations are presented in Table 6. The highest frequency was found in Lithuanians, followed by two subpopulations of Latvians, Estonians, Belarusians, and Russians. It follows from the frequency distribution that *LW^b* frequency displayed a steady continuous decrease trend in all directions away from the central area of the Balts. The *LW^b* allele is found also in other Slavic-speaking populations, in particular the Poles. This fact may reflect the peculiarities of the ethnic history of Slavs: the presence of *LW^b* allele may reflect the admixture with the Baltic-speaking populations, if the mutation originated from the Balts, or revealing a trace of an ancient autochthonous community that existed in the territory of East-European plain up to the Baltic Sea. The peculiar geographical distribution of *LW^b* allele, with relatively high frequencies in the Balts and zero or almost zero levels in Asiatic, African, and West-European populations (The 1000 Genomes Project), makes this blood group a very informative marker for Baltic genetic influence. This means that its presence in other populations is likely to be due to Baltic admixture.

Gene *MCM6*, allele *13910*T*. Gene *MCM6* (OMIM *601806) encodes protein minichromosome maintenance complex component 6, which participates in the replication of eukaryotic DNA. Gene *MCM6* in humans is located upstream from the lactase gene *LCT* (OMIM *603202) and regulates its transcription. Gene *LCT* encodes enzyme lactase, which catalyses the hydrolysis of lactose, the main carbohydrate in milk, to monosaccharide glucose and galactose. Lactase activity in new-born mammals is high. In adult mammals, including humans, lactase persistence (LP) or lactose tolerance normally disappears after weaning. LP is an autosomal dominant trait enabling continued produc-

tion of lactase throughout adult life. Lactase non-persistence (lactose intolerance, lactase deficiency, adult type hypolactasia), which is an ancestral condition for all mammals, is an autosomal recessive trait (OMIM #223100). The degree and symptoms of lactose intolerance vary individually, and most lactose intolerant individuals can usually tolerate small amounts of lactose. The geographic distribution of LP matches the distribution of dairy farming: LP is common in North-European populations (increase from southeast to northwest across Europe), as well as in some pastoralist populations from Africa and the Middle East (Scrimshaw and Murray, 1988; Laland *et al.*, 2010). In European populations, two dominant alleles in the gene *MCM6* that directly affect *LCT* gene promoter activity and are tightly associated with LP are found: -22018*A (rs182549) in intron 9, and -13910*T (rs4988235) in intron 13 (Enattah *et al.*, 2002). In Europeans the best studied is the association of LP with the genotypes C/T and T/T in the above-mentioned locus -13910. Frequencies of *MCM6* gene allele -13910*T in Latvians and other European populations as well as frequencies of LP are shown in Table 7. The highest frequencies are observed in North Western and Northern Europe: in Irish, Swedes, Finns, Danes, and English. It is estimated that allele -13910*T derived from the ancestral allele -13910*C during the past 6100–7500 years within Central Europeans, probably in response to selection for the ability to digest milk as adults (Bersaglieri *et al.*, 2004). Allelic age estimates are consistent with the results of an ancient DNA study, which showed that allele -13910*T was rare or absent among early farmers from Central and Eastern Europe (Burger *et al.*, 2007). The observed high frequencies of allele -13910*T in Northern Europe can best be explained by strong selection that took place in Neolithic Central Europe and was followed by a migration of people representing this culture to Northern Europe (Vuorisalo *et al.*, 2012). The immigrants gradually replaced or were assimilated by the local hunter-gatherer populations. The low frequencies of *MCM6* gene -13910*T allele and lactase persistence in Latvians and North-Western Russians may be explained by the relatively recent transition from Mesolithic to Neolithic in the Baltic region, which was not driven by admixture with early European farmers (Jones *et al.*, 2017). The analysis of ancient DNA has shown that the transition from Mesolithic to Neolithic is also the period when first derived *MCM6* allele, which is associated with lactase persistence, is found in Latvia. There is increased frequency of the derived alleles in the Bronze Age samples (0.50 in 9 Latvian individuals) (Table 1, Jones *et al.*, 2017; Mitnik *et al.*, 2017a; 2017b).

MITOCHONDRIAL DNA VARIATION

Mitochondrial DNA (mtDNA) – matrilineal – genetic variation was investigated by Pliss *et al.*, 2006 in a sample of 299 (in her PhD thesis — 351) healthy unrelated Latvians representing the above-mentioned four different regions of Latvia. Sequencing of the first mtDNA hypervariable segment (HVS-I) in combination with analysis of informative coding region markers revealed that the vast majority of the

Table 6

FREQUENCIES OF GENE *ICAM4* ALLELE LW^B (RS77493670) IN EUROPEAN POPULATIONS

Populations	n	gene <i>ICAM4</i> allele LW ^B (rs77493670)
Latvians ¹		
NW	84	0.05
CE	67	0.02
SW	36	0.03
E	75	0.06
Total	262	0.04
Lithuania ¹		
Žemaitija	84	0.07
East Aukštaitija	38	0.05
West Aukštaitija	29	0.04
South Aukštaitija	34	0.03
Total	185	0.06
Russians (Vologda) ²	383	0.02
Belarusians ¹	60	0.03
Poles ²	749	0.02
Swedes (Central) ²	395	0.01
English people ³	91	0.01
Italians ³	100	0
Estonians ²	800	0.04
Finns ²	908	0.03

¹ Gailite (Piekuse), unpublished data; ² Sistonen *et al.*, 1999; ³ 1000 Genomes Browser

observed mtDNA variants in Latvians belong to haplogroups (hgs) como most European populations (Table 8). One hundred and twenty four different haplotypes (nucleotide sequence variants within haplogroups) belonging to 11 major hgs were observed. HgH, which is the most frequent hg in European populations, accounted for almost half (43.3%) of mtDNA variants in Latvians. 45% of the classified hg H cases belonged to sub-hgs H1 and H5. Sub-hg H1b, which occurs more frequently in Eastern and Central North Europe, was the most abundant type of sub-hg H1 among Latvians, also being frequent among Estonians. The HVS-I sequence motifs characteristic to this sub-hg can often be observed also in the Lithuanian population (Kasperavičiūtė *et al.*, 2004). In the Latvian gene pool, H1b was significantly more frequent than among Eastern Slavs (Loogväli *et al.*, 2004). Significant differences of H1a frequencies were observed between Latvians and Finns (the statistical significance of population differences for mtDNA data is taken from Pliss *et al.*, 2006). Among the Finns, sub-hg H1b was not observed. On the other hand, sub-hg H1f, frequent in the Finnish population, was not found in the Latvian mtDNA pool. Sub-hg H5, the second largest in the Latvian mtDNA pool (15% of H), was found to be significantly more frequent among Latvians than in the gene pool of hg H in Eastern Slavs. A quarter of mtDNA variants belonged to hg U. Five sub-hgs (U2, U3, U4, U5 and U8) were observed among hg U haplotypes. The sub-hgs U4 and U5 were most abundant, covering about 40% of the hg U

Table 7

FREQUENCIES OF GENE *MCM6* ALLELE -13910*T IN EUROPEAN POPULATIONS

Populations	n	<i>MCM6</i> gene -13910*T allele frequency (rs4988235)	Lactase persistence frequency ⁶
Latvians ¹			
NW	86	0.37	nda
CE	67	0.32	nda
SW	56	0.34	nda
E	68	0.30	nda
Total	277	0.34	0.55
Lithuanians		0.32 ⁹	0.61
Russians (NW) ¹			
Staritsa, Tver Oblast	21	0.38	nda
Sychevka, Smolensk Oblast	29	0.33	nda
Ustyuzhna, Vologda Oblast	40	0.36	nda
Mezen, Archangelsk Oblast	32	0.44	nda
Total	122	0.38	
Russians (Central)	50	0.24 ²	0.43
Belarusians		nda	0.62
Ukrainians		nda	0.54
Poles	289	0.46 ⁷	0.63
Czechs		nda	0.82
Irish	65	0.95 ³	0.96
Swedes (Central, S)	360	0.82 ²	0.96
Norwegians		nda	0.96
Danes	97811	0.76 ⁸	0.96
English people	3344	0.75 ⁴	0.95
Germans	60	0.56 ³	0.85
Italians (North)	28	0.36 ²	0.49
Italians (Middle)	16	0.06 ²	nda
French	58	0.43 ²	0.71
Estonians	355	0.51 ⁵	0.64
Finns ²	360	0.82 ²	0.84

nda, no data available; ¹ Gailite (Piekuse), unpublished data; ² Bersaglieri *et al.*, 2004; ³ Itan *et al.*, 2009; ⁴ Smith *et al.*, 2009; ⁵ Lember *et al.*, 2006; ⁶ Cook, 2014; ⁷ Madry *et al.*, 2011; ⁸ Bergholdt *et al.*, 2015; ⁹ Kalibatiene *et al.*, 2006

gene pool. Comparison of mtDNA hg profiles in different European populations (Table 7) showed that the frequency of hg U4 was significantly higher in Eastern than in Western European populations. Interestingly, the frequency of sub-hg U4 in Latvians is among the highest in Europe. In the central part of Latvia (Semigalia), U4 was found in 14.9% of all mtDNA variants. Pliss *et al.* (2006) noted that the heterogeneity of hg U4 is also relatively high in Semigalia: of thirteen U4 haplotypes found among Latvians eight were discovered in this region. Another example of only few region-specific differences in the Latvian mtDNA pool is provided by sub-hg U2, which has a particularly high frequency in Eastern Latvia; seven of nine Latvian sub-hg U2 mtDNAs were found in Latgale. The frequency of U2 was significantly higher in Latgale than in the western parts of

MITOCHONDRIAL DNA HAPLOGROUP FREQUENCIES (%) AMONG LATVIANS AND OTHER EUROPEAN POPULATIONS

Haplogroups	Latvians ¹					Lithuanians ²	Russians ³	Belarusians ⁴	Ukrainians ⁵	Poles ³	Germans ⁶	Estonians ⁶	Finns ⁶	Mari ⁷
	NW	CE	SW	E	Total									
R0 (xHVO, H)	1.1	1.5	2.9	2.3	2.0	1.7	2.5	4.5	0.0	0.9	0.9	1.5	0.0	1.5
HV0	3.4	3.0	1.5	3.1	2.8	6.1	5.5	5.2	1.7	4.8	4.5	3.9	6.8	11.0
H	54.5	46.3	38.2	36.7	43.3	46.1	42.3	36.7	41.3	45.2	47.7	43.5	40.5	40.4
U (total)	11.4	32.8	23.5	30.5	24.8	17.8	17.9	20.2	19.1	16.0	13.5	24.2	24.9	24.3
U1	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.3	0.2	0.0	0.0
U2	0.0	3.0	0.0	5.5	2.6	0.0	1.5	1.5	1.7	0.9	0.9	1.5	0.7	0.0
U3	0.0	3.0	2.9	1.6	1.7	1.7	1.0	2.3	0.6	0.5	1.2	1.0	0.0	0.0
U4	6.8	14.9	10.3	4.7	8.3	5.0	3.5	3.8	5.0	5.0	1.8	5.4	1.0	10.3
U5	4.5	12.0	10.3	17.2	11.7	8.9	10.4	12.0	11.8	8.7	8.4	14.2	22.4	14.0
U5a	4.5	9.0	7.4	10.2	8.0	3.9	7.5	6.4	5.6	5.3	nda	9.8	8.0	nda
U5b	0.0	3.0	2.9	7.0	3.7	5.0	2.9	5.6	6.2	3.4	nda	4.4	14.4	nda
U6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
U7	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.4	0.0	0.2	0.3	0.0	0.3	0.0
U8	0.0	0.0	0.0	0.8	0.3	1.1	0.0	0.0	0.0	0.5	0.0	1.7	0.5	0.0
U*(other U)	0.0	0.0	0.0	0.8	0.3	1.1	0.0	0.4	0.0	0.2	0.6	0.2	0.0	0.0
K	0.0	1.5	1.5	4.7	2.3	2.2	3.0	3.8	6.7	3.4	7.5	2.7	2.9	2.2
J	6.8	6.0	10.3	4.7	6.6	7.8	8.0	9.0	6.2	7.8	8.4	10.3	4.8	7.4
T	11.4	4.5	13.2	8.6	9.4	10.0	10.9	7.5	12.9	11.5	9.0	7.8	3.8	5.1
N1 (xI)	0.0	0.0	0.0	0.0	0.0	1.1	0.0	3.4	1.7	0.4	0.6	1.5	0.0	0.0
I	5.7	3.0	4.4	5.5	4.8	3.9	2.5	2.6	2.8	1.8	1.8	1.0	4.1	0.7
W	5.7	1.5	4.4	2.3	3.4	1.1	2.0	3.4	1.7	3.7	2.7	2.4	8.8	0.0
X	0.0	0.0	0.0	0.8	0.3	1.1	3.5	1.1	1.7	1.8	1.2	1.0	0.9	0.0
M	0.0	0.0	0.0	0.8	0.3	0.0	1.5	2.6	1.1	1.8	0.9	0.2	1.9	5.9
others	0.0	0.0	0.0	0.0	0.0	1.1	0.5	0.5	3.3	0.7	1.2	0.0	0.5	1.5
n	88	67	68	128	351	180	201	267	179	436	333	409	580	136

nda, no data available; ¹ Pliss, 2007; ² Kasperavičiūte *et al.*, 2004; ³ Malyarchuk *et al.*, 2002; ⁴ Kushniarevich *et al.*, 2013; ⁵ Kushniarevich *et al.*, 2015; ⁶ data from Pliss *et al.*, 2006; ⁷ Bermisheva *et al.*, 2002

Latvia (Northern and Southern Curonia). The authors suggested that the high frequency of hg U2 in Latgale could be best explained by a recent founder effect, because all U2 samples found belonged to the same HVS-I haplotype. Similar intra-population homogeneity was also observed among Lithuanians (Kasperavičiūte *et al.*, 2004). U5a, another frequent haplogroup (8.0% in Latvians), is more typical for Eastern Europeans compared to Central and South-Eastern Europe, whereas haplogroup U5b (3.7% in Latvians) reflects the input from South-Western and Central Europe (Malyarchuk *et al.*, 2010). The other hg that has a different frequency pattern in Eastern and Western European populations (Table 8) is K; statistically significant differences of hg K frequencies were observed between Latvians and Western European populations, with hg K being more common among the latter. Only a single member of Asian-specific hg M (G2a) was found in Latgale, the eastern part of Latvia. In Northern Europe, the two main hg M lineage groups are D5b and Z1, although they are present in many populations at low frequencies (Tambets *et al.*, 2004).

Interestingly, these hgs were observed neither in the sample of 351 Latvians, nor among 225 Lithuanians (Kasperavičiūte *et al.*, 2004; Tambets *et al.*, 2004). One D5b individual has been found among 545 Estonians (Tambets *et al.*, 2004). This observation is of some interest because the northward Finnic-speaking populations, Finns, Saami, and Karelians, possess East Asian maternal lineages at frequencies that are low, but still higher (2%–5%) (Sajantila *et al.*, 1995; Tambets *et al.*, 2004) than in the populations inhabiting the South-Eastern Baltic coast. Pliss *et al.* (2006) suggested that East Asian maternal gene flow had, in Eastern Europe, only a negligible impact on populations living in the South-Eastern Baltic Sea region, although it reached populations living close to the sub-Arctic fringe of Europe.

To illustrate the genetic relationships of studied populations, principal component (PC) analysis based on the frequencies of mtDNA hgs from Table 8 was performed (Fig. 6). The analysis showed that all compared European populations, except Finns and Mari, fall into one cluster. Latvi-

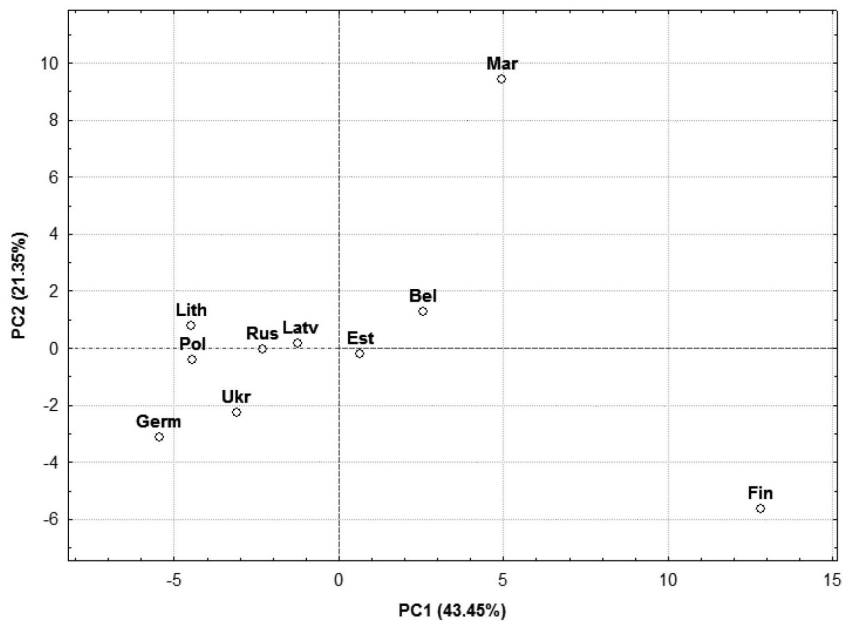


Fig. 6. Principal component analysis (PCA) based on mtDNA haplogroup frequencies in some European populations. Population codes are given in alphabetical order as follows: Bel, Belarusians; Est, Estonians; Fin, Finns; Ger, Germans; Lat, Latvians; Lit, Lithuanians; Mar, Maris; Pol, Poles; Rus, Russians; Ukr, Ukrainians. The genetic variation retained by different components is shown in brackets.

ans and Lithuanians formed a relatively tight cluster with Estonians, Russians, Poles, and Belarusians.

The HVS-I haplotype-sharing analysis performed by Pliss *et al.* (2006) among Baltic-, Germanic-, Slavic-, and Finno-Ugric speaking populations showed that the vast majority of mtDNA haplotypes found among Latvians are identical to, or close derivatives of, those observed in other Eastern and Western European populations, irrespective of their linguistic affiliations. These results most likely reflect a deep common origin for the European mtDNA pool (Richards *et al.* 1996). It might be that no significant differences in mtDNA variability between founder populations existed in the period of formation of the Balts; therefore one cannot distinguish different components of founders in the present day populations.

The recent development of high-throughput sequencing of mtDNA genomes has revealed some matrilineal gene pool substructure in Europe which was not detected using only HVS1 sequences and low resolution haplogroup information (Torron *et al.*, 2006). In order to obtain important additional information on modern Latvian mtDNA pool, future analysis of high resolution data is needed.

It seems important that in recent analysis of twenty-one ancient Mesolithic and Neolithic DNA samples from the Zvejnieki burial site in Northern Latvia, all of them belonged to typically hunter-gatherer haplogroup U: eleven to sub-lineage U5a, two — U5b, seven — U4, one — U2 (Jones *et al.*, 2017, Mathieson *et al.*, 2017; Table 1). Haplogroup U is found in most European hunter-gatherer samples (Skoglund *et al.*, 2012; Brandt *et al.*, 2013; Posth *et al.*, 2016). No mitochondrial haplogroups that are typical for Central European Mesolithic/Neolithic farmers were detected in the territory of today's Latvia. The Bronze Age individuals from Ķivutkalns burial display a more diverse set of mitochondrial haplogroups, besides U, including also H (different sublineages), J and T, which first appeared in

Europe during the Neolithic (Mittnik *et al.*, 2017a; 2017b; Table 1).

NRY VARIATION

Several studies of Y-chromosomal – patrilineal – variability in the Latvian population have been conducted, based on biallelic markers and short tandem repeats (STRs). Contrary to mtDNA data, analysis of Y-chromosome lineages has revealed deeper differences between the studied North-Eastern European populations. Rosser *et al.* (2000) on analysis of 11 biallelic NRY polymorphisms in 47 European populations (including Latvians) concluded that Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Based on the analysis of both Y-chromosome haplogroup frequency distribution and microsatellite variation, Zerjal *et al.* (2001) hypothesised that Y chromosomes of Finno-Ugric- and Baltic-speaking populations have distinct genetic histories. These results were confirmed by Lessig *et al.* (2001) and Kasperavičiūtė *et al.* (2004). They noted, however, that it is unclear whether the observed differences suggested different source populations or rather more recent random genetic drift. The analysis of five biallelic Y-chromosomal markers performed by Laitinen *et al.* (2002), on the other hand, demonstrated a high genetic similarity between Estonian, Latvian, Lithuanian, and Finno-Ugric-speaking Mari males, leading to the suggestion that Baltic males share common Finno-Ugric-speaking forefathers (Laitinen *et al.*, 2002). The genotyping of DNA samples subsequently performed by Lappalainen *et al.* (2008) on a larger group of markers (18 Y-chromosomal biallelic polymorphisms and 9 STRs) confirmed differences in the haplotype structure of Baltic-speaking Latvians and Lithuanians and Finno-Ugric populations, allowing to suggest at least two main migration waves to the Baltic Sea region (Lappalainen *et al.*, 2008). Kushniarevich *et al.* (2015) in analysis of 40 binary NRY markers found that both extant Baltic-speaking populations, Latvi-

NRY HAPLOGROUP FREQUENCIES (%) IN LATVIANS AND COMPARABLE EUROPEAN POPULATIONS

Haplogroup	Markers	Latvians ¹					Lithuanians ²	Russians Northern ³	Russians Central ³	Belarusians ⁴	Ukrainians ⁵	Poles ⁶	Germans ⁶	Estonians ²	Finns ²	Mari ⁷
		NW	CE	SW	E	Total										
DE	M145/YAP	3.5	0	3.0	1.7	1.9	1.2	0.2	5.0	3.4	8.2	2.9	6.6	2.5	0.4	0
F*	M89 (x G, H, I, J, K)	0	0	0	1.7	0.6	1.8	0.8	0.8	0.4	4.8	0	0	0.8	0.4	0
G	M201	0	0	0	1.7	0.6	0	1.2	0	1.4	0	1.5	2.5	0	0	0
I (total)	M170	13.8	10.3	3.0	10.3	9.4	11.6	13.1	17.5	23.4	28.8	17.1	21.0	16.9	28.9	8.1
I*	M170 (xII, I2a, I2b)	0	0	0	0	0	0	0.8	0.6	0	nda	0	0	0	0	nda
II	M253	13.8	7.7	3.0	3.5	6.3	4.9	6.2	5.3	5.0	nda	10.0	14.8	11.9	28.0	nda
I2a	P37	0	0	0	1.7	0.6	4.9	5.7	10.0	17.2	nda	5.4	3.0	4.2	0.2	nda
I2b	M223	0	2.6	0	5.1	2.5	1.8	0.4	1.6	1.2	nda	1.7	3.2	0.8	0.7	nda
R*	M207 (xR1a, R1b, R2)	0	0	0	0	0	0	0	0	0.2	0	0	0	0	0.2	0
R1a	SRY1532	44.8	38.5	39.4	32.8	37.7	34.1	34.2	46.5	50.6	44.5	60.6	26.5	37.3	7.1	47.7
R1b	M343	6.9	7.7	15.2	3.5	7.6	4.9	5.4	7.5	6.0	8.2	11.9	36.7	4.2	3.7	2.7
Q	M242	0	0	3.0	0	0.6	0	0.7	0.3	0.2	0.2	0.6	0.2	0.8	0.2	0
N*	M231 (xN3, N2a)	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0.2	0
N2a (prev. N1b)	P43	0	0	0	0	0	0	7.5	0.5	0.5	0	0	0	0	0.4	9.9
N3 (prev. N1c)	M46/Tat	31.0	43.6	36.4	48.3	41.5	43.9	35.5	16.3	9.6	0	3.1	1.1	33.9	58.2	31.5
J	M304	0	0	0	0	0	1.8	1.8	3.4	3.9	3.4	2.3	5.1	1.7	0	0
K*	M9 (xN, P, R, T, O, L)	0	0	0	0	0	0	0	1.7	0	2.1	0	0	0.8	0	0
P*	P295 (Q, R)	0	0	0	0	0	0.6	0	0	0	0	0	0	0.8	0	0
Y*	SRY10831a (xA,B,C,D, E,F,P)	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0
C	M130	0	0	0	0	0	0	0	0.2	0	0	0	0.2	0	0.2	0
T	M70	0	0	0	0	0	0	0	0	0.5	0	0	0.6	0	0	0
n		29	39	33	58	159	164	380	364	565	146	520	472	118	536	111

¹ Pliss *et al.*, 2015; ² Lappalainen *et al.*, 2008; ³ Balanovsky *et al.*, 2008; ⁴ Kushniarevich *et al.*, 2013; ⁵ Mielnik-Sikorska *et al.*, 2013; ⁶ Rebała *et al.*, 2012; ⁷ Tambets *et al.*, 2004

ans and Lithuanians, lie in the vicinity of Finno-Ugric speaking Estonians according to their Y-chromosome diversity. It is also worth mentioning that the population genetic analysis on samples from North-Western Russia indicated a discrepancy in the depth of penetration of mtDNA and Y-chromosome lineages characteristic for the majority of South-Western Russian populations in the east and north of Eastern Europe (Malyarchuk *et al.*, 2004). These results led the authors to conclude that Russian colonisation of north-eastern European territories might have been accomplished mainly by males rather than by females.

The most detailed analysis of NRY in Latvians was performed by Pliss *et al.* (2015), who investigated 159 unrelated ethnic Latvian males from the above-mentioned four

different geographic regions, using 28 biallelic markers and 12 STRs (Pliss *et al.*, 2015). Eleven different haplogroups (hgs) (Table 9) and 130 distinctive haplotypes were detected in a regionally relatively homogeneous Latvian population. The vast majority of the analysed samples belonged to two predominant hgs, N3 (formerly N1c) and R1a (41.5% and 37.7%, accordingly). Twelve samples (7.6%) represented hg R1b, which is carried predominantly by Western Europeans (Eastern Europeans, in contrast, have high frequency of R1a lineages). The next abundant variant among Latvians was hg II with a frequency of 6.3%. Another two phylogenetic branches of hg I occur at a low level in Latvians: hg I2b was found at a frequency of 2.5% and hg I2a comprised only 0.6%. In general, the gene pool of the Latvian Y-chromosomes was characterised by the pres-

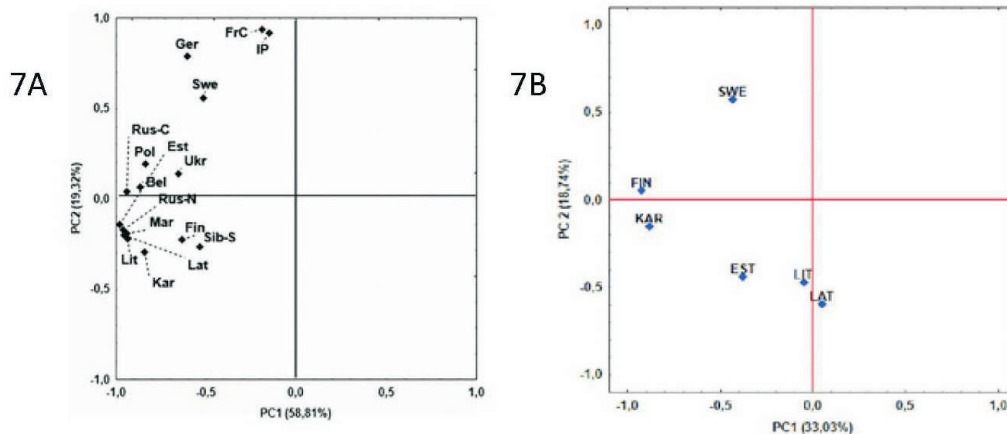


Fig. 7. Comparison of Latvian population with other populations based on variation of the Y-chromosomal haplogroups (7A) and compound haplotypes (7B) (Pliss *et al.*, 2015). 7A, population codes are given in alphabetical order as follows: Bel, Belarusians; Est, Estonians; Fin, Finns; FrC, French and Corsicans; Ger, Germans; IB, Iberian Peninsula's populations; Kar, Karelians; Lat, Latvians; Lit, Lithuanians; Mar, Maris; Pol, Poles; Rus N/C, Russians Northern/Central; Sib-S, South Siberians; Swe, Swedes; Ukr, Ukrainians. PC1 and PC2, principal component 1 and 2 are significant.

7B, population codes are given in alphabetical order as follows: EST, Estonians; FIN, Finns; KAR, Karelians; LAT, Latvians; LIT, Lithuanians; SWE, Swedes. PC1 and PC2, principal component 1 and 2 are significant.

ence of the majority of hgs typical to Eastern and Northern Europe, and only one sample fell into hg Q (0.6%), which is more characteristic to Asian populations, but is occasionally also found in Europe (Tambets *et al.*, 2004; Balanovsky *et al.*, 2008; Lappalainen *et al.*, 2008). Statistically significant differences in the distribution of hgs between the sampled Latvian subpopulations were not found. Results of principal component analysis (PCA) performed on the distribution of hgs among 16 populations revealed that the Latvian population formed a tight cluster with the North-European populations, irrespective of linguistic affiliation, that is, with Estonians, Maris, Northern Russians, and Lithuanians, all having a high frequency of N3 (Fig. 7A). For the analysis of compound haplotypes, nine YSTRs were used, and only six European populations were included due to the absence of a comparable set of Y-STR data in other neighbouring populations. Results of PCA grouped Baltic-speaking Latvians and Lithuanians together and lightly split them from Finno-Ugric-speaking Estonians (Fig. 7B). Such an observation could be explained by different haplotype composition of the most abundant Y-chromosomal lineages, namely, N3 and R1a, in those populations. The majority of haplotypes (24) shared between Latvians and neighbouring populations were found for hg R1a. Only four compound haplotypes were in common within hg N3. Despite the comparable frequency distribution of hg N3 in Latvians and Lithuanians with the Finno-Ugric-speaking populations from the eastern coast of the Baltic Sea, the observed differences in allelic variances of N3 haplotypes between these two groups are in concordance with the previously stated hypothesis (Zerjal *et al.*, 2001; Kasperavičiūtė *et al.*, 2004; Lappalainen *et al.*, 2008) of two migration waves that introduced hg N3 founder haplotypes to the eastern coast of the Baltic Sea: “Baltic” — in the south of the Baltic Sea through the Baltic countries to Poland and Germany; and “Fennoscandian” — in the north of the Baltic Sea through Finland to Sweden. It was suggested that N3 chromosomes originated in Asia and

were brought to Europe by Finno-Ugric speakers prior to the arrival of Indo-European speakers. Recently performed high-coverage Y chromosome sequencing (a method that allows to identify previously undetected lineages) from 56 Eurasian populations, including two Latvian samples, showed that both Latvian samples belong to haplogroup N3, clade N3a3-VL29 (Karmin *et al.*, 2015; Ilumäe *et al.*, 2016). This clade is found in over a third of the present-day male Estonians, Latvians, and Lithuanians, and it is present also among Finns, Karelians, Saami. It was suggested that the distribution of this Y chromosome lineage has been shaped by random genetic drift in small populations dispersed across a wide area (Ilumäe *et al.*, 2016).

Hg R1a is frequent in North-Eastern Europe, with the highest frequencies (30%–50%) in Slavic-speaking populations and moderate among populations of Volga-Ural basin. Recently, Underhill *et al.* (2015) underwent a detailed phylogeographic study of European-specific R1a sub-clades of R1a-M458 and R1a-M558 lineages. R1a-M458 shows the highest distribution among European frequencies (exceeding 30%) in Central and Eastern Europe (Underhill *et al.*, 2010). The highest frequency of R1a-M458 was found in Poles, Czechs, Slovaks, and Western Belarusians (20%). Unlike R1a-M458, the R1a-M558 lineage is more common in Eastern Slavs and in Volga-Uralic populations. R1a-M558 is found in various parts of Russia (10%–30%), as well as in Poland and Western Belarus where it transcends 26%, and in Ukraine, where its frequency varies between 10% and 23%, whereas it occurs at significantly lower frequencies in Western Europe (Underhill *et al.*, 2015). The study of R1a Y-chromosomes in the Latvian sample performed by Pliss *et al.* (2015) showed that the R1a-M458 lineage was found at a relatively low frequency, as opposed to the R1a-M558 lineage, which accounts for nearly all R1a hg representatives and more than a third of all Latvian paternal lineages. The authors suggested that high prevalence

of R1a-M558 in Latvians and their haplotypes shared with Eastern Slavic-speaking populations may indicate a shared genetic ancestry component or a possible more recent gene flow from Eastern Slavic-speaking populations to present-day Latvians. There were many R1a-M558 haplotypes in Latvians that were detected also among Eastern Slavic-speaking Russians, Ukrainians, and Belarusians (Underhill *et al.*, 2015).

Recent analysis of fourteen Mesolithic and Neolithic ancient male DNA samples from the archaeological site of Zvejnieki in Northern Latvia showed haplogroups R1b (eight samples), I2a (five samples) and Q (one sample) (Table 1; Jones *et al.*, 2017; Mathieson *et al.*, 2017). R1b, the most common hg found in modern Western Europeans, has been found at low frequencies before the Late Neolithic in Western Europe but at higher frequencies in Russia. It was suggested that this hg spread into Western Europe from the East after 5000 cal YBP (Haak *et al.*, 2015). Haplogroup I2a has been commonly found among Western hunter-gatherers (Lazaridis *et al.*, 2014). The analysis of six Bronze Age individuals from Kivutkalns burial revealed haplogroups R1a (five individuals) and R1b (one individual) (Mittnik *et al.*, 2017a, b). No cases of haplogroup N, which is a very common Y chromosome haplogroup in modern Latvians (41.5%), Lithuanians, Finns and Estonians (Ilumäe *et al.*, 2016) were detected among Latvian ancient DNA samples. It is proposed that this haplogroup was brought to the Eastern Baltic after the Late Bronze Age (Mittnik *et al.*, 2017a).

CONCLUSIONS

Population genetic studies of Latvians, which represent one of the two extant Baltic-speaking populations besides Lithuanians, have shown a mixture of genetic traits typical for Western Europe and Eastern Eurasia with only small differences between Latvian subpopulations.

The analyses of genome-wide single nucleotide polymorphisms have shown that Latvians and Lithuanians clustered in close proximity to Russians from the central part of European Russia and Estonians, although Estonians had a higher proportion of ancestry component common with Finns.

The studies of the Baltic “tribal gene” LW^b as well as gene’s *SERPINA1* allele PIZ indicate the presence of a considerable Baltic admixture in the neighbouring Finno-Ugric and Slavic populations.

The analyses of mitochondrial DNA have shown that in general Latvians belong to the same common gene pool as most Europeans, with Latvians and Lithuanians forming a tight cluster with Estonians, Russians, Belarusians, and Poles. The vast majority of mtDNA HVS-I haplotypes found among Latvians are identical to or close derivatives of those observed in other Eastern and Western European populations, irrespective of their linguistic affiliations.

The analyses of Y-chromosome lineages, contrary to mtDNA data, have revealed deeper differences between North-European populations, with Lithuanians, Estonians, and Eastern-Slavic populations being ethnogenetically closest to Latvians.

Studies of ancient DNA samples from the territory of today’s Latvia have shown some hunter-gatherer ancestry persistence across the Neolithic transition, impact from Northern Eurasian populations between the Early Neolithic and the Middle Neolithic as well as impact from the Pontic Steppe populations during the Late Neolithic and Bronze Age. There are no signs of Anatolian farmer-related genetic input on the population of Latvia at least until the end of the Bronze Age. Further research of ancient DNA samples from different archaeological cultures with comparison to present-day populations with detailed joint analysis of genetic, craniometric, archaeological, and linguistic data is needed to reveal demographic events that created present-day Latvian genetic diversity.

ACKNOWLEDGEMENTS

The authors are grateful to Prof. Elmārs Grēns (Latvian Biomedical Research and Study Centre) for support and encouragement, Prof. Andrejs Vasks (Institute of Latvian History, University of Latvia) and Jānis Ciglis (National History Museum of Latvia) for helpful information on the history of Latvia; to all individuals who donated their genomic material for the studies.

REFERENCES

- Ajioka, R. S., Jorde, L. B., Gruen, J. R., Yu, P., Dimitrova, D., Barrow, J., Radisky, E., Edwards, C. Q., Griffen, L. M., Kushner, J. P. (1997). Haplotype analysis of hemochromatosis: Evaluation of different linkage-disequilibrium approaches and evolution of disease chromosomes. *Amer. J. Hum. Genet.*, **60**, 1439–1447.
- Allentoft, M. E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., Schroeder, H., Ahlström, T., Vinner, L., Malaspina, A.-P., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P. D., Dąbrowski, P., Duffy, P. R., Ebel, A. V., Epimakhov, A., Frei, K., Furmanek, M., Gralak, T., Gromov, A., Gronkiewicz, S., Grupe, G., Hajdu, T., Jarysz, R., Khartanovich, V., Khokhlov, A., Kiss, V., Kolij, J., Kriiska, A., Lasak, I., Longhi, C., McGlynn, G., Merkevicius, A., Merkyte, I., Metspalu, M., Mkrtychyan, R., Moiseyev, V., Paja, L., Pálfi, G., Pokutta, D., Pospieszny, L., Price, T. D., Saag, L., Sablin, M., Shishlina, N., Vjčlav Smrčka, V., Soenov, V. I., Szeverényi, V., Tóth, G., Trifanova, S. V., Varul, L., Vicze, M., Yepiskoposyan, L., Zhitenev, V., Orlando, L., Sichevitz-Pontén, T., Brunak, S., Nielsen, R., Kristiansen, K., Eske Willerslev, E. (2015). Population genomics of Bronze Age Eurasia. *Nature*, **522**, 167–172.
- Balanovsky, O., Pocheshkhova, E., Pshenichnov, A., Solovieva, D., Kuznetsova, M., Voronko, O., Churnosov, M., Tegako, O., Atramentova, L., Lavryashina, M., Evseeva, I., Borinska, S., Boldyreva, M., Dubova, N., Balanovska, E. (2005). Is spatial distribution of the HIV-resistant CCR5Δ32 allele formed by ecological factors? *J. Physiol. Anthropol. Appl. Hum. Sci.*, **24**, 375–382.
- Balanovsky, O., Rootsi, S., Pshenichnov, A., Kivisild, T., Churnosov, M., Evseeva, I., Pocheshkhova, E., Boldyreva, M., Yankovsky, N., Balanovska, E., Villems, R. (2008). Two sources of the Russian patrilial heritage in their Eurasian context. *Amer. J. Hum. Genet.*, **82**, 236–250.

- Beckman, L., Sikstrom, C., Mikelsaar, A.-V., Krumina, A., Kučinskas, V., Beckman, G. (1999). α_1 -antitrypsin (PI) alleles as markers of West-European influence in the Baltic Sea region. *Human Heredity*, **49**, 52–55.
- Bergholdt, H. K. M., Nordestgaard, B. G., Varbo, A., Ellervik, C. (2015). Milk intake is not associated with ischaemic heart disease in observational or Mendelian randomization analyses in 98529 Danish adults. *Int. J. Epidemiol.*, **44** (2), 587–603.
- Bermisheva, M., Tambets, K., VILLEMS, R., Khusnutdinova, E. (2002). Diversity of mitochondrial DNA haplotypes in ethnic populations of the Volga-Ural region of Russia. *Mol. Biol. (Moscow)*, **36**, 802–812.
- Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., Rhodes, M., Reich, D. E., Hirschhorn, J. N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *Amer. J. Hum. Genet.*, **74**, 1111–1120.
- Bērziņš, V., Vasks, A. (2013). Latvijas aizvēsture (līdz 800. gadam) [Prehistory of Latvia (until 800)]. In: Stradiņš, J. (Ed.). *Latvieši un Latvija* [Latvians and Latvia], 1 sēj. *Latvieši* [Latvians]. LZA, Rīga, pp. 51–83 (in Latvian).
- Blanco, I., de Serres, F.J., Fernandes-Bustillo, E., Lara, B., Miravittles, M. (2006). Estimated numbers and prevalence of PI S and PI Z alleles of α_1 -antitrypsin deficiency in European countries. *Eur. Respir. J.*, **27**, 77–84.
- Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., Gray, R. D., Suchard, M. A., Atkinson, Q. D. (2012). Mapping the origins and expansion of the Indo-European language family. *Science*, **337**, 957–960.
- Brandt, G., Haak, W., Adler, C. J., Roth, C., Szécsényi-Nagy, A., Karimnia, S., Möller-Rieker, S., Meller, H., Ganslmeier, R., Friederich, S., Dresely, V., Nicklisch, N., Pickrell, J. K., Sirocko, F., Reich, D., Cooper, A., Alt, K.W., Genographic Consortium. (2013). Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science*, **342**, 257–261.
- Burger, J., Kirchner, M., Bramanti, B., Haak, W., Thomas, M.G. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc. Natl. Acad. Sci. USA*, **104**, 3736–3741.
- Byrne, K. M., Byrne, P. C. (2004). Review: Other blood group systems – Diego, Yt, Xg, Scianna, Dombrock, Colton, Landsteiner-Wiener, and Indian. *Immunohematology*, **20**, 50–58.
- Carroll, T. P., O'Connor, C. A., Floyd, O., McPartlin, J., Kelleher, D. P., O'Brien, G., Dimitrov, B. D., Morris, V. B., Taggart, C. C., McElvaney, N. G. (2011). The prevalence of alpha-1 antitrypsin deficiency in Ireland. *Resp. Res.*, **12**, 91–98.
- Cauna, N. (1942). Par latviešu asins grupām [On Latvian blood groups]. *Ārstniecības žurnāls*, Nr. 1, 254–260 (in Latvian).
- Cook, C. J. (2014). The role of lactase persistence in precolonial development. *J. Econ. Growth*, **19**, 369–406.
- Česnys, G. (1991). Anthropological roots of Lithuanians. *Science, Arts Lith.*, **1**, 4–10.
- Davydenko, O. G., Kushniarevich, A. I., Sivitskaya, L. N., Danilenko, N. G. (2007). Ethnogenomics of Belarusians. Preliminary results [Давыденко, О. Г., Кушнеревич, Е. И., Сивецкая, Л. Н., Даниленко, Н. Г. Этногеномика белорусского народа. Предварительные итоги]. *Molecular Appl. Gen. [Молекулярная и прикладная генетика]*, **6**, 45–78 (in Russian).
- de Juan, D., Reta, A., Castiella, A., Pozueta, J., Prada, A., Cuadrado, E. (2001). *HFE* gene mutations analysis in Basque hereditary haemochromatosis patients and controls. *Eur. J. Hum. Genet.*, **9**, 961–964.
- Denisova, R. (1975). *Anthropology of Ancient Balts* [Денисова, Р. А. Антропология древних балтов]. Zinātne, Rīga. 400 pp. (in Russian).
- Denisova, R. (1997). Latvijas senākā apdzīvotība [The initial settlement of Latvia]. *J. Latv. Hist. Inst.*, No. 3, 5–15 (in Latvian).
- de Silva, E., Stumpf, M. P. H. (2004). HIV and the CCR5-Delta32 resistance allele. *FEMS Microbiol. Lett.*, **241**, 1–12.
- Distante, S., Robson, K. J., Graham-Campbell, J., Arnaiz-Villena, A., Brissot, P., Worwood, M. (2004). The origin and spread of the *HFE*-C282Y haemochromatosis mutation. *Hum. Genet.*, **115**, 269–279.
- Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L., Järvelä, I. (2002). Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.*, **30**, 233–237.
- Farrell, P. M. (2008). The prevalence of cystic fibrosis in the European Union. *J. Cystic Fibr.*, **7**, 450–453.
- Faure, E., Royer-Carenzi, M. (2008). Is the European spatial distribution of the HIV-resistant CCR5- Δ 32 allele formed by a breakdown of the pathocenosis due to the historical Roman expansion? *Infect. Genet. Evol.*, **8**, 864–874.
- Galvani, A. P., Novembre, J. (2005). The evolutionary history of the CCR5-Delta32 HIV-resistance mutation. *Microbes Infect.*, **7**, 302–309.
- Galvani, A. P., Slatkin, M. (2003). Evaluating plague and smallpox as historical selective pressures for the CCR5-delta 32 HIV-resistance allele. *Proc. Natl. Acad. Sci. USA*, **100**, 15276–15279.
- Gimbutiene (Gimbutas), M. (1994). *Balti aizvēsturiskajos laikos* [Balts in prehistoric times]. Zinātne, Rīga. 226 pp. (in Latvian).
- Gray, R. D., Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, **426**, 435–439.
- Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Binffy, E., Economou, C., Francken, M., Friederich, S., Pena, R. G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S. L., Risch, R., Rojo Guerra, M. A., Roth, C., Szécsényi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K. W., Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, **522** (7555), 207–211.
- Hakulinen, L. (1961). *The Structure and Development of the Finnish Language*. Curzon. 383 pp.
- Hummel, S., Schmidt, D., Kremeyer, B., Herrmann, B., Oppermann, M. (2005). Detection of the CCR5-Delta32 HIV resistance gene in Bronze Age skeletons. *Genes Immun.*, **6**, 371–374.
- Hutchison, D. C. (1998). Alpha-1-antitrypsin deficiency in Europe: Geographical distribution of Pi types S and Z. *Resp. Med.*, **92**, 367–377.
- Ilumäe, A.-M., Reidla, M., Chukhryaeva, M., Järve, M., Post, H., Karmin, M., Saag, L., Agdzhoyan, A., Kushniarevich, A., Litvinov, S., Ekomasova, N., Tambets, K., Metspalu, E., Khusainova, R., Yunusbayev, B., Khusnutdinova, E. K., Osipova, L. P., Fedorova, S., Utevska, O., Koshel, S., Balanovska, E., Behar, D. M., Balanovsky, O., Kivisild, T., Underhill, P. A., VILLEMS, R., Rootsi, S. (2016). Human Y chromosome haplogroup N: A non-trivial time-resolved phylogeography that cuts across language families. *Amer. J. Hum. Genet.*, **99**, 163–173.
- Itan, Y., Powell, A., Beaumont, M. A., Burger, J., Thomas, M. G. (2009). The origins of lactase persistence in Europe. *PLoS Computational Biology*, **5**, e1000491.
- Ivanova, A., von Ahsen, N., Adjarov, D., Krastev, Z., Oellerich, M., Wieland, E. (1999). C282Y and H63D mutations in the *HFE* gene are not associated with porphyria cutanea tarda in Bulgaria. *Hepatology*, **30**, 1531–1532.
- Jones, E. R., Zarina, G., Moiseyev, V., Lightfoot, E., Nigst, P. R., Manica, A., Pinhasi, R., Bradley, D. G. (2017). The Neolithic transition in the Baltic was not driven by admixture with early European farmers. *Curr. Biol.*, **27**, 576–582.
- Kalibatiene, D., Varvuolyte, S., Alekna, V. (2006). Investigation of hypolactasia in the elderly. *Gerontologija*, **7**, 136–142.

- Pliss, L., Timša, L., Rootsi, S., Tambets, K., Pelnen, I., Zole, E., Puzuka, A., Sabule, A., Rozane, S., Lace, B., Kucinskas, V., Krumina, A., Ranka, R., Baumanis, V. (2015). Y-chromosomal lineages of Latvians in the context of the genetic variation of the Eastern-Baltic region. *Ann. Hum. Genet.*, **79**, 418–430.
- Posth, C., Renaud, G., Mittnik, A., Drucker, D. G., Rougier, H., Cupillard, C., Valentin, F., Thevenet, C., Furtwängler, A., Wißing, C., Francken, M., Malina, M., Bolus, M., Lari, M., Gigli, E., Capocchi, G., Crevecoeur, I., Beauval, C., Flas, D., Germonpré, M., van der Plicht, J., Cottiaux, R., Gély, B., Ronchitelli, A., Wehrberger, K., Grigorescu, D., Svoboda, J., Semal, P., Caramelli, D., Bocherens, H., Harvati, K., Conard, N. J., Haak, W., Powell, A., Krause, J. (2016). Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a Late Glacial population turnover in Europe. *Curr. Biol.*, **26**, 827–833.
- Prīmanis, J. (1937). *Latviešu antropoloģiskais raksturojums* [Anthropological Characterization of the Latvians]. Valtera un Rapas akc. sab.apgāds, Rīga. 40 lpp. (in Latvian).
- Proņina, N., Lugojska, R. (2011). Association between minihaplotypes and mutations at the phenylalanine hydroxylase locus in Latvian phenylketonuria patients. *Proceed. Latv. Acad. Sci., Section B*, **65** (3/4), 73–79.
- Purina, G., Lugojska, R., Sokolova, L. (1995). Medical genetical service in Latvia: Developmental trends. *Proceed. Latv. Acad. Sci.*, No. 5/6, 5–8.
- Radiņš, A. (2012). *Arheoloģisks ceļvedis latviešu un Latvijas vēsturē* [Archaeological Guide in the History of Latvians and Latvia]. Apgāds Neputns, Rīga. 415 lpp. (in Latvian).
- Rebala, K., Martinez-Cruz, B., Tönjes, A., Kovacs, P., Stumvoll, M., Lindner, I., Büttner, A., Wichmann, H.-E., Siváková, D., Soták, M., Quintana-Murci, L., Szczerkowska, Z., Comas, D., Genographic Consortium (2012). Contemporary paternal genetic landscape of Polish and German populations: From early medieval Slavic expansion to post-World War II resettlements. *Eur. J. Hum. Genet.*, **21**, 415–422.
- Richards, M., Corte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H.-J., Sykes, B. (1996). Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Amer. J. Hum. Genet.*, **59**, 185–203.
- Rimantiene, R. (1996). *Akmens amžius Lietuvoje* [The Stone Age in Lithuania]. 2nd edn. 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., Kučinskas, 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., Previderè, 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. *Amer. J. Hum. Genet.*, **67**, 1526–1543.
- Saag, L., Varul, L., Scheib, C. L., Stenderup, J., Allentoft, M. E., Saag, L., Pagani, L., Reidla, M., Tambets, K., Metspalu, E., Kriiska, A., Willerslev, E., Kivisild, T., Metspalu, M. (2017). Extensive farming in Estonia started through a sex-biased migration from the Steppe. *Curr. Biol.*, **27**, 2185–2193.
- Sajantila, A., Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M. L., Aula, P., Beckman, L., Tranebjaerg, L., Gedde-Dahl, T., Issel-Tarver, L., Di Rienzo, A., Paabo, S. (1995). Genes and languages in Europe: An analysis of mitochondrial lineages. *Genome Res.*, **5**, 42–52.
- Scrimshaw, N., Murray, E. (1988). The acceptability of milk and milk products in populations with a high prevalence of lactose intolerance. *Amer. J. Clin. Nutr.*, **48**, 1079–1159.
- Sedovs V. (2004). *Balti senatnē* [Balts in Ancient Times]. 2nd ed. Latvijas Vēstures Institūta apgāds, Rīga. 165 lpp. (in Latvian).
- Seixas, S., Garcia, O., Trovada, M. J., Santos, M. T., Amorim, A., Rosha, J. (2001). Patterns of haplotype diversity within the serpin gene cluster at 14q32.1: Insights into the natural history of the alpha 1-antitrypsin polymorphism. *Hum. Genet.*, **108**, 20–30.
- Sistonen, P., Virtaranta-Knowles, K., Denisova, R., Kucinskas, V., Ambra-siene, D., Beckman, L. (1999). The LWb blood group as a marker of prehistoric Baltic migrations and admixture. *Hum. Hered.*, **49**, 154–158.
- Sivitskaya, L., Danilenko, N., Zabarouskaya, Z., Davydenko, O. (2013). *HFE* gene mutation associated with the severity of gestational diabetes mellitus in Belarussian women. *Open J. Endocr. Metab. Dis.*, **3**, 13–17.
- Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M. T., Götherström, A., Jakobsson, M. (2012). Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science*, **336**, 466–469.
- Slatkin, M. (2001). Simulating genealogies of selected alleles in a population of variable size. *Genet. Res.*, **78**, 49–57.
- Smith, G. D., Lawlor, D. A., Timpson, N. J., Baban, J., Kiessling, M., Day, I. N. M., Ebrahim, S. (2009). Lactase persistence-related genetic variant: Population substructure and health outcomes. *Eur. J. Hum. Gen.*, **17**, 357–367.
- Srivastava, K., Almarry, N. S., Flegel, W. A. (2014). Genetic variation of the whole *ICAM4* gene in Caucasians and African Americans. *Transfusion*, **54**, 2315–2324.
- Stepanova, A. A., Krasovsky, S. A., Polyakov, A. V. (2016). Reliability of the search for 19 common mutations in the CFTR gene in Russian cystic fibrosis patients and the calculated frequency of the disease in Russian Federation. *Rus. J. Genet.*, **52**, 204–213.
- Stephens, J. C., Reich, D. E., Goldstein, D. B., Shin, H. D., Smith, M. W., Carrington, M., Winkler, C., Huttley, G., Allikmets, R., Schriml, L., Gerrard, B., Malasky, M., Ramos, M. D., Morlot, S., Tzetis, M., Oddoux, C., Giovine, F. S., Nasioulas, G., Chandler, D., Aseev, M., Hanson, M., Kalaydjieva, L., Glavac, D., Gasparini, P., Kanavakis, E., Claustres, M., Kambouris, M., Ostrer, H., Duff, G., Baranov, V., Sibul, H., Metspalu, A., Goldman, D., Martin, N., Duffy, D., Schmidtke, J., Estivill, X., O'Brien, S. J., Dean, M. (1998). Dating the origin of the CCR5-Delta 32 AIDS-resistance allele by the coalescence of haplotypes. *Amer. J. Hum. Genet.*, **62**, 1507–1515.
- Stumpf, M. P., Wilkinson-Herbots, H. M. (2004). Allelic histories: Positive selection on a HIV-resistance allele. *Trends Ecol. Evol.*, **19**, 166–168.
- Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogväli, E.-L., Tolk, H.-V., Reidla, M., Metspalu, E., Pliss, L., Balanovsky, O., Pshenichnov, A., Balanovska, E., Gubina, M., Zhadanov, S., Osipova, L., Damba, L., Voevoda, M., Kutuev, I., Bermisheva, M., Khusnutdinova, E., Gusar, V., Grechanina, E., Parik, J., Pennarun, E., Richard, C., Chaventre, A., Moisan, J.-P., Barac, L., Peričić, M., Rudan, P., Terzić, R., Mikerezi, I., Krumina, A., Baumanis, V., Koziel, S., Rickards, O., De Stefano, G. F., Anagnou, N., Pappa, K. I., Michalodimitrakis, E., Ferjk, V., Füredi, S., Komel, R., Beckman, L., Villems, R. (2004). The Western and Eastern roots of the Saami — the story of genetic “outliers” told by mitochondrial DNA and Y chromosomes. *Amer. J. Hum. Genet.*, **74**, 661–682.
- Tighe, O., Dunican, D., O'Neill, C., Bertorelle, G., Beattie, D., Graham, C., Zschocke, J., Cali, F., Romano, V., Hrabincova, E., Kozak, L., Nechyporenko, M., Livshits, L., Guldberg, P., Jurkowska, M., Zekanowski, C., Perez, B., Desviat, L. R., Ugarte, M., Kucinskas, V., Knappskog, P., Treacy, E., Naughten, E., Tyfield, L., Byck, S., Scriver, C. R., Mayne, P. D., Croke, D. T. (2003). Genetic diversity within the R408W Phenylketonuria mutation lineages in Europe. *Hum. Mutat.*, **21**, 387–393.
- Torroni, A., Achilli, A., Macalay, V., Richards, M., Bandelt, H. J. Harvesting the fruit of the human mtDNA tree. *Trends Genet.*, **22** (6), 339–345.
- Underhill, P. A., Myres, N. M., Rootsi, S., Metspalu, M., Zhivotovsky, L. A., King, R. J., Lin, A. A., Chow, C. E., Semino, O., Battaglia, V., Kutuev, I.,

- Järve, M., Chaubey, G., Ayub, Q., Mohyuddin, A., Mehdi, S. Q., Sengupta, S., Rogaev, E. I., Khusnutdinova, E. K., Pshenichnov, A., Balanovsky, O., Balanovska, E., Jeran, N., Augustin, D. H., Baldovic, M., Herrera, R. J., Thangaraj, K., Singh, V., Singh, L., Majumder, P., Rudan, P., Primorac, D., Vilems, R., Kivisild, T. (2010). Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur. J. Hum. Genet.*, **18**, 479–484.
- Underhill, P. A., Poznik, G. D., Rootsi, S., Järve, M., Lin, A. A., Wang, J., Passarelli, B., Kanbar, J., Myres, N. M., King, R. J., di Cristofaro, J., Sahakyan, H., Behar, D. M., Kushniarevich, A., Šarac, J., Šaric, T., Rudan, P., Pathak, A. K., Chaubey, G., Grugni, V., Semino, O., Yepiskoposyan, L., Bahmanimehr, A., Farjadian, S., Balanovsky, O., Khusnutdinova, E. K., Herrera, R. J., Chiaroni, J., Bustamante, C. D., Quake, S. R., Kivisild, T., Vilems, R. (2015). The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur. J. Hum. Genet.*, **23**, 124–131.
- Vasks, A., Vaska, B., Gravere, R. (1997). *Latvijas aizvēsture* [The Prehistory of Latvia]. Zvaigzne ABC, Rīga, 223 lpp. (in Latvian).
- Voicu, P. M., Cojocariu, C., Petrescu-Danila, E., Covic, M., Stanciu, C., Rusu, M. (2009). Prevalence of *HFE* (hemochromatosis) gene mutations C282Y and H63D in Romanian population. *Blood Cells Mol. Dis.*, **42**, 14–15.
- Vuorisalo, T., Arjamaa, O., Vasemägi, A., Taavitsainen, J.-P., Tourunen, A., Saloniemi, I. (2012). High lactose tolerance in North Europeans: A result of migration, not *in situ* milk consumption. *Persp. Biol. Med.*, **55**, 163–174.
- Williams, R. A., Mamotte, C. D. S., Burnett, J. R. (2008). Phenylketonuria: An inborn error of phenylalanine metabolism. *Clin. Biochem.*, **29**, 31–41.
- World Health Organization. The molecular genetic epidemiology of cystic fibrosis. Available at: www.who.int/genomics/publications/en/HGN_WB_04.02_report.pdf
- Zagorska, I. (2012). *Senie ziemeļbriežu mednieki Latvijā* [The Ancient Reindeer Hunters in Latvia]. Zinātne, Rīga. 206 lpp. (in Latvian).
- 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.

ONLINE RESOURCES

- 1000 genomes browser: phase3browser.1000genomes.org (accessed 15.12.2017).
- dbSNP database: <https://www.ncbi.nlm.nih.gov/snp/> (accessed 15.12.2017).
- Ensembl database: <https://www.ensembl.org> (accessed 15.12.2017).
- ExAc database: exac.broadinstitute.org/ (accessed 15.12.2017).
- Online Mendelian Inheritance of man: <https://www.omim.org> (accessed 15.12.2017).

Received 24 July 2017

Accepted in the final form 2 January 2018

LATVIEŠU POPULĀCIJU ĢENĒTIKA SAISTĪBĀ AR ZIEMEĻAUSTRUMEIROPAS ETNISKO GRUPU SAJAUKUMU

Pārskata rakstā ir apkopoti dati par populāciju ģenētiskajiem pētījumiem latviešiem — vienai no divām mūsdienās baltu valodās runājošajām populācijām. Rakstā sniegta informācija par genoma viena nukleotīda polimorfismu plašas analīzes datiem, dots salīdzinājums ar apkārtējo populāciju datiem, izmantojot arī klasisko autosomālo marķieru, mitochondriālās DNS (mtDNS), Y hromosomas nerekombinējošās daļas, kā arī no seno kaulu paraugiem izdalītās DNS pētījumu rezultātus. Latviešiem ir vērojama Rietumeiropas un Austrumeiropas populācijām raksturīgo ģenētisko pazīmju kombinācija, atklājot tikai nelielas atšķirības starp mūsdienu Latvijas reģioniem. Baltu izcelsmes alēles LWb (*ICAM4* gēnā) un gēna *SERPINA1* alēles PIZ pētījumi norāda uz vērā ņemamu baltu ģenētiskā materiāla klātbūtni kaimiņu somugru un slāvu populācijās. Lai gan mtDNS pētījumi liecina, ka latvieši ģenētiski ir līdzīgi lielākajai daļai eiropiešu populāciju, Y hromosomas analīzes dati norāda, ka vislielākā līdzība latviešiem ir vērojama ar Ziemeļeiropas un Austrumeiropas kaimiņu populācijām — lietuviešiem, igauņiem un austrumslāviem. Senās DNS analīze no Agrā un Vidējā Neolīta mednieku un vācēju kaulu paraugiem, kas iegūti mūsdienu Latvijas un Lietuvas teritorijā, neatklāj Centrāleiropas zemkopju genomu piejaukuma pēdas.