MITOCHONDRIAL DNA (mtDNA) VARIATION AND ITS ROLE IN ETHNOGENESIS OF LATVIANS

Liana Pliss

Latvian Biomedical Research and Study Centre

UNIVERSITY OF LATVIA

RIGA 2007

UNIVERSITY OF LATVIA

Liana Pliss

MITOCHONDRIAL DNA (mtDNA) VARIATION AND ITS ROLE IN ETHNOGENESIS OF LATVIANS

Author:……Liana Pliss

Promoter:……Viesturs Baumanis
MD, Professor, Dr. biol.

Latvian Biomedical Research and Study Centre

RIGA 2007

University of Latvia
Promotion Council
Biology
Riga, Latvia

Chairman:…….Pauls Pumpens
Professor, Dr. habil. biol.

Opponents: **Dr. habil. biol., Prof.** **Elmars Grens**
Latvian Biomedical Research and Study Centre, Riga, Latvia

Dr. habil. biol. **Izaks Rasals**
LU Institute of Biology, Salaspils, Latvia

Dr. habil. med., Prof. **Vaidutis Kucinskas**
Vilnius University, Vilnius, Lithuania

Dissertation is accepted for the commencement of the degree of Dr. biol. (in Molecular Biology)

Commencement: June 8th, 2007

Pliss L. 2007. Mitohondriālās DNS (mtDNS) polimorfismi latviešu etnoģenēzes pētījumos. Latvijas Universitāte, Rīga, 75 lpp.

KOPSAVILKUMS

Cilvēka mitohondriju genomu (mtDNS) šodien plaši lieto, lai aprakstītu cilvēka populācijas izcelsmi un migrācijas virzienus. Mitohondriālās DNS variāciju kartēšana Āzijā, Amerikā, Āfrikā un Eiropā dod ieskatu par cilvēka populācijas izplatīšanās virzieniem un izkliedi pasaulē. mtDNS īpatnības (maternālā pārmantošana, rekombināciju trūkums) ļauj pēc reģistrētajām mutācijām izsekot cilvēces vēstures ģenealoģiju. Šī ģenealoģija dod iespēju analizēt cilvēka populācijas vēsturi, pirmsvēsturi un prognozēt to attīstību. Liela mēroga Eiropas populāciju mtDNS daudzveidības pētījumi galvenokārt tika veikti tās Rietumu un Centrālās daļās. Tomēr, pēdējo gadu laikā sāka parādīties dati par mitohondriālo variantu dažādību Austrumu Eiropas un slāvu valodās runājošās populācijās. Līdz šim laikam, divās baltu valodās runājošās tautas – latviešos un lietuviešos - populāciju ģenētikas pētījumi tika veltīti pārsvarā tikai autosomālo jeb klasisko ģenētisko marķieru (piemēram, TF*DCH1, PI, LWb) analīzēm un Y hromosomas polimorfismu variācijām. Tādēļ, Latvijas populācijas mitohondriju ģenētiskā heterogenitāte, salīdzinot to ar citu Eiropas populāciju datiem, ir aktuāls ģenētisko pētījumu virziens.

Šī darba galvenais mērķis bija atrast un analizēt iedzimto mtDNS variāciju sastopamību Latvijas populācijā un tās etnolingvistiskajās grupās, kā arī veikt daudzpusīgo filoģeogrāfisko analīzi inter- un intra – populāciju līmenī, lai izprastu mtDNS polimorfismu ģenētiskās daudzveidības izcelsmi un to raksturu Latvijas populācijā. Tādā veidā novērota maternālo līniju ģenētiskā struktūra var sniegt informāciju par Latvijas populācijas etnoģenēzi. Šai darbā tika veikta mtDNS variāciju analīze Latvijas populācijā un tās etnolingvistiskajās grupās (analizēto paraugu skaits – 351), pielietojot molekulārās bioloģijas metodes, statistiskās un filoģenētiskās analīzes. Populāciju salīdzinājuma analīzē tika izmantota dažādu Eirāzijas populāciju plašā mtDNS datu bāze (publicēto 11,236 tūkstošu paraugu un nepublicēto 4,732 tūkstošu paraugu mtDNS datu bāze). Balstoties uz vairāku analīžu datiem, iegūtā mtDNS variāciju izplatības biežuma sastopamība dažādos Latvijas reģionos tika salīdzināta ar lingvistisko un antropoloģisko pētījumu rezultātiem. Konstatēts mtDNS ģenētisko variantu iekšējais migrācijas virziens no Latvijas austrumu/dienvidaustrumu daļas (Latgales) uz ziemeļrietumu un centrālo daļu, un noskaidrotais ģenētisko variāciju plūsmas virziens ir saskaņā ar eksistējošiem arheoloģiskajiem datiem. Tika veikta Latvijas populācijā atrasto mtDNS variāciju plaša mēroga filoģeogrāfiskā analīze, kā arī aprakstīta ģenētiskā saistība ar citām Eiropas populācijām (gan indoeiropiešu valodās runājošajām, gan somu – ugru valodās runājošajām populācijām). Parādīts, ka, balstoties uz mtDNS „haplotipu – līdzības" analīzes rezultātiem starp baltu, ģermāņu, slāvu, un somu – ugru valodās runājošajām populācijām, lielākā novēroto mtDNS haplotipu daļa Latvijas populācijā ir līdzīga vai pat identiska ar konstatētajiem mtDNS variantiem citās Austrumu un Rietumu Eiropas populācijās neatkarīgi no populāciju ģeogrāfiskās vai lingvistiskās piederības. Iegūtie rezultāti liecina par kopīgu Eiropas populācijās eksistējošo mtDNS variantu izcelsmi.

Izveidots priekšstats par Latvijas populācijā esošo mtDNS ģenētisko variabilitāti var kalpot kā papildus informācijas avots turpmākajiem ģenētisko asociāciju pētījumiem: dažādu fenotipisko izpausmju saistību ar mtDNS polimorfismiem, kā arī būs iespējams veikt ar mitohondriālo iedzimtību saistīto ģenētisko slimību (optiskā neiropātija, miopātija, encefalomiopātija u.c.) diagnostiku, kas Latvijā pagaidām netiek veikta. mtDNS mutāciju detektēšana dažādās vecuma grupās ļaus noskaidrot mitohondriālā genoma polimorfismu lomu novecošanās, kā arī to saistību ar mtDNS replikāciju un transkripciju.

Promocijas darbs ir izstrādāts laika periodā no 2002. līdz 2007. gadam Latvijas Biomedicīnas Pētījumu un Studiju Centrā, zinātniskā vadītāja LU Prof. Viestura Baumaņa vadītā

molekulārās mikrobioloģijas laboratorijā sadarbībā ar RSU Medicīnas Bioloģijas un Ģenētikas katedru (Prof. Astrīda Krūmiņa), Tartu Universitātes Evolucionārās Bioloģijas katedru (Prof. Richard Villems, Tartu, Igaunija), Bernes Universitātes, Populāciju ģenētikas katedru (Prof. Laurent Excoffier, Berne, Šveice) un Imunoloģijas Institūtu (Prof. Artūrs Sočnevs). Pētījumu atbalsta Eiropas Struktūras Fonda (ESF) programma (ESS2004/3) un Latvijas Zinātnes Padomes sadarbības projekts (01.0023).

Pliss L. 2007. Mitochondrial DNA (mtDNA) variation and its role in ethnogenesis of Latvians. 75 pp.

SUMMARY

Variation in the human mitochondrial genome is now routinely described and used to infer the origin and migration patterns in human populations as well as range of expansions and gene flow during the prehistory. Mapping of the variation in the modern mtDNA landscape of Asia, America and Europe is shedding the light on the patterns of colonization and dispersal of peoples in the world. Due to the mode of inheritance and mutation rate of mtDNA the mutations that have struck throughout human history trace the maternal genealogy. This genealogy can be used in attempt to make inferences about prehistoric processes of human populations. Large-scale studies of European mtDNA diversity have been mostly concentrated on Western and Central European populations. In the last few years, data on maternal lineages from Eastern European and Slavonic-speaking populations have started to emerge. So far, population genetic studies of the two extant Baltic-speaking populations – Latvians and Lithuanians, who form a separate branch of the Indo-European language family tree – have mostly touched on the variation of autosomal or classical genetic markers (*e.g.* TF*DCH1, PI, LWb) and the pattern of spread of Y-chromosomal variation. Therefore, genetic differentiation within the Latvian population and the relationships between it and other European populations are aspects that require further detailed study.

The main goal of the present study was to obtain the knowledge about mtDNA variation in Latvia and its sub-populations, and further to perform a large-scale phylogeographic analysis at inter- and intra-population level in order to understand the origin and rise of the genetic diversity of Latvian mtDNA lineages. The analysis of the mtDNA variation in the Latvian population as a whole and it subpopulations (sample size – 351 individuals) has been performed using methods of molecular biology, statistical and phylogenetic approaches. For comparison, the mtDNA database of 11,236 individuals, consisting of published mtDNA data as well as unpublished mtDNA data from 4,732 individuals from different Eurasian populations were used as background information for the analysis. Based on results of the different analyses observed mtDNA variation distribution in Latvian regions was compared with linguistic and anthropological research data. The genetic trace of the forced internal migration movement from eastern/south-eastern region (Lettigalia) towards the north-western and central parts of the country can be seen based on mtDNA variation, and is in accordance with archaeological data. In order to find out genetic relationship with other European populations (Indo-European- and Finno-Ugric-speaking populations) the large-scale phylogeographic analysis of the revealed mtDNA variation distribution in Latvian population was undertaken. The HVS-I haplotype-sharing analysis among Baltic-, Germanic-, Slavonic-, and Finno-Ugric-speaking populations have 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 the linguistic affiliations of the latter. Pattern of mtDNA lineages and its variation distribution across Eastern European populations indicate on or most likely reflect an ancient and largely common heritage of European populations.

Established notion of the Latvian mtDNA genetic variability further can be used in case-control association studies: possible connection of different phenotypical manifestations with mitochondrial variability in Latvian population (inherited mitochondrial disorders, *e.g.* Leber's hereditary optic neuropathy (LHON) and chronic progressive external ophtalmoplegia (CPEO). In the future we can implement the association study on mtDNA variation spread and its molecular characteristics in two age groups (control group and elderly individuals).

The present study was carried out during the time period from 2002 to 2007 in the Laboratory of Molecular Microbiology (led by LU Prof. Viesturs Baumanis), Latvian Biomedical Research and Study Centre in collaboration with Department of Medical Biology and Genetics (Riga Stradins University, Prof. Astrida Krumina), Department of Evolutionary Biology (University of Tartu, Estonia, Prof. Richard Villems), Department of Population Genetics (University of Berne, Switzerland, Prof. Laurent Ecoffier) and Institute of Immunology (Prof. Arturs Socnevs). This research was supported by a grant from the European Social Fund (ESF) program ESS2004/3 and Latvian Council of Science National program 01.0023.

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The current dissertation is based on the following publication referred to in the text by their Roman numerals:

I. **Pliss, L.,** Tambets, K., Loogväli, E.-L., Pronina, N., Lazdins, M., Krumina, A., Baumanis, V., Villems, R. (2006) Mitochondrial DNA portrait of Latvians: towards the understanding of the genetic structure of Baltic-speaking populations. *Annals of Human Genetics* 70, 439-458.

II. Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogväli, E.-L., Tolk, H.-V., Reidla, M., Metspalu, E., **Pliss, L.,** 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., Pericic, M., Rudan, P., Terzic, R., Mikerezi, I., Krumina, A., Baumanis, V., Koziel, S., Rickards, O., De Stefano, G.-F., Anagnou, N., Pappa, K.I., Michalodimitrakis, E., Ferak, V., Furedi, S., Komel, R., Beckman, L., Villems, R. (2004) The western and eastern roots of the Saami – the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *American Journal of Human Genetics* 74, 661-682.

III. Timsa, L., Puzuka, A., **Pliss, L.,** Lazdins, M., Baumanis, V., Krumina, A. (2005) Human Y chromosome and its role in human pathology and population phylogenetic studies. *Proceedings of the Latvian Academy of Sciences* 59, 93-107.

IV. Pronina, N., **Pliss, L.,** Lazdins, M., Baumanis, V., Krumina, A. (2003) The mitochondrial genome and human population studies. *Proceedings of the Latvian Academy of Sciences* 57, 199-208.

My contribution to the articles referred in the current thesis is as follows:

Ref. **I** – a) conceived and designed the mtDNA experiments; b) performed the mtDNA experiments of the Latvian sample, analyzed the mtDNA variation of the populations listed in *Subjects and Methods*; c) performed the statistical and phylogeographical analysis of mtDNA; d) wrote the paper;

Ref. **II** – a) participated in the performing the experiments: analyzed mtDNA variation (HVS-I sequencing, RFLP analysis and sequencing of the informative coding sites of 31 U5 genomes of Latvians; b) assisted in the analysis of the data; c) in the preparation of the manuscript;

Ref. **III** – a) corresponding literature overview; b) in the preparation of the manuscript;

Ref. **IV** – a) corresponding literature overview; b) wrote the paper;

ABBREVIATIONS

AFLP	amplified fragment length polymorphisms
AMH	Anatomically modern human
AMOVA	analysis of molecular variance
bp/kbp	base pairs/thousand (kilo) base pairs
CRS	Cambridge Reference Sequence
D-loop	displacement loop/control region of mtDNA
DNA	deoxyribonucleic acid
hg(s)	haplogroup(s)
HVS-I/HVS-II	the first/second hypervariable segment
KY	thousands years
KYA	thousands (kilo) years ago
LD	Linkage Disequilibrium
LGM	the Last Glacial Maximum
LW^b	Landsteiner-Wiener (LW) blood group (*LW* – gene, b – allele)
MRCA	the most recent common ancestor
MSY	male-specific region of the Y chromosome
mtDNA	mitochondrial DNA
MYA	million years ago
NJ	neighbour joining
np(s)	nucleotide position(s)
nDNA	nuclear DNA
OXPHOS	oxidative phosphorylation
PCA	principal component analysis
PCR	Polymerase Chain Reaction
PI	alpha 1-antitrypsin (PI) alleles
PKU	phenylketonuria
RFLP	Restriction Fragment Length Polymorphism
RNA	ribonucleic acid
SD	standard deviation
SNP	single nucleotide polymorphism
rRNA	ribosomal ribonucleic acid
TF (DCH1)	transferrin gene (TF) (DCH1 – allele)
tRNA	transfer ribonucleic acid
TMRCA	time to most recent common ancestor
YBP	years before present

Definitions of basic terms used in the current dissertation

Haplotype (=lineage)	mtDNA sequence with characteristic polymorphisms, encompasses all identical sequences;
Haplogroup	monophyletic clade of haplotypes sharing characteristic defining sequence polymorphisms;
Founder haplotype	common ancestral haplotype to which all haplotypes under concern coalesce to;
Coalescent time	time to MRCA;
Phylogeography	the study of the spatial distribution of genealogical lineages;
Star-like phylogeny	phylogeny of a set of sequences that mostly (or all) coalesce to the same haplotype
Pleistocene	a geological epoch that lays between about 2 MYA and 10 KYA
Holocene	the last ~11 KY, with an unusually warm and stable climate

1. INTRODUCTION

Since the first studies that showed considerable mitochondrial DNA (mtDNA) variation in different geographical subsets of humans, the analysis of mtDNA polymorphisms has been a favoured tool in population genetics. In uniparentally inherited non-recombining human mtDNA, polymorphisms have accumulated sequentially along radiating maternal lineages from sets of often continent-specific mtDNA founders, providing a detailed record of the ancient migration patterns of women. Large-scale studies of European mtDNA diversity have been mostly concentrated on Western and Central European populations. In the last few years, data on maternal lineages from Eastern European and Slavonic-speaking populations have started to emerge. So far, population genetic studies of the two extant Baltic-speaking populations – Latvians and Lithuanians, who form a separate branch of the Indo-European language family tree – have mostly touched on the variation of classical genetic markers and the pattern of spread of polymorphisms associated with diseases, as well as variation of the Y chromosome. Recently, it has been showed that mtDNA variation among different Lithuanian regions is limited and that Lithuanian mtDNA variation is closely related both to Slavonic- and Finno-Ugric-speaking populations of Northern and Eastern Europe. Genetic investigations of autosomal or classical genetic markers (*e.g.* TF DCH1, PI, and LW^b) among Latvians and Lithuanians have revealed genetic stratification of Baltic-speakers at the intra-population level, as well as differences in the Baltic-speakers compared to other Indo-European and Finno-Ugric-speaking populations of the Baltic Sea region. The analysis of one of the most common genetic diseases in Europeans – phenylketonuria (PKU) – has revealed the predominance of a single mutation in the phenylalanine hydroxylase (*PAH*) gene, R408W of haplotype 2, in the Baltic States. Although well spread throughout Eastern Europe, R408W has its frequency peak in Latvia, Lithuania and Estonia, there comprising about four fifth of PKU haplotypes. It has been suggested that this mutation originated in an ancient eastern European population, from where it spread westward. Meanwhile, the analysis of Y-chromosomal markers showed the highest genetic similarities between Estonian, Latvian and Lithuanian males, followed by Finno-Ugric-speaking Mari. However, later it has been concluded, based on the analysis of both haplogroup frequencies distribution and microsatellite variation, that the Y chromosomes of Finno- Ugric- and Baltic-speaking populations have distinct genetic histories. One of the main players in Y chromosomal variation in Northern Eurasian populations is haplogroup N3 (formerly hg 16 or TatC allele). This haplogroup is frequent among Finno-Ugric-speaking populations and many Siberian populations, but present only at very low frequencies in Southern and Western Europe. Haplogroup N3 is also frequent in Baltic-speaking populations, but less so in the Slavonic-speaking neighbours of Latvians and Lithuanians, even though the two linguistic families – Slavonic and Baltic – are sister groups in the Indo- European tree of languages.

The first section of this study gives an overview about the special properties of mtDNA and its worldwide variation, with an emphasis on the European mtDNA variety. It also includes a short description of methods used in phylogenetic analyses, and an overview of current hypotheses on the origin of anatomically modern humans (AMHs). Knowledge about the general topology of the global mtDNA tree provides the basis for investigating many interesting details of mtDNA variation in different regions.

The main aim of the research, described in the results and discussion part of this thesis, was to improve our understanding of the general processes that have shaped the landscape of mtDNA variation of the present-day Latvian population.

2. LITERATURE OVERVIEW

2.1. Structure and organization of human mitochondrial DNA (mtDNA)

Constitutively non-recombining region of our genetic material is mitochondrial DNA (mtDNA), a circular double-stranded DNA molecule about 16.5 kb in length whose entire sequence is known (Anderson *et al.,* 1981; Andrews *et al.,* 1999). This is contained not within nucleus, but within mitochondria – cytoplasmic organelles in which the energy-generating process takes place. Each mitochondrion contains usually 5-10 mtDNAs in its matrix. The number of mtDNAs in somatic cell is about 1,000-10,000 (Lightowlers *et al.,* 1997). As mitochondria contain ribosomes and DNA, and are only formed by the division of other mitochondria, it is generally accepted that they were originally derived from endosymbiotic prokaryotes. Studies of mitochondrial DNA, which is often circular and employs a variant genetic code, show their ancestor, the so-called proto-mitochondrion, was a member of the Proteobacteria (Futuyma and Douglas, 2005). In particular, the pre-mitochondrion was probably related to the rickettsias, although the exact position of the ancestor of mitochondria among the alpha-proteobacteria remains controversial. The endosymbiotic hypothesis suggests that mitochondria descended from specialized bacteria (probably purple non-sulfur bacteria) that somehow survived endocytosis by another species of prokaryote or some other cell type, and became incorporated into the cytoplasm. The ability of symbiont bacteria to conduct cellular respiration in host cells that had relied on glycolysis and fermentation would have provided a considerable evolutionary advantage. Similarly, host cells with symbiotic bacteria capable of photosynthesis would also have an advantage. In both cases, the number of environments in which the cells could survive would have been greatly expanded. This relationship developed at least 2 billion years ago and mitochondria still show some signs of their ancient origin. Mitochondrial ribosomes are the 70S (bacterial) type, in contrast to the 80S ribosomes found elsewhere in the cell. As in prokaryotes, there is a very high proportion of coding DNA, and an absence of repeats. Mitochondrial genes are transcribed as multigenic transcripts which are cleaved and polyadenylated to yield mature mRNAs. Unlike their nuclear cousins, mitochondrial genes are small, generally lacking introns, and many chromosomes are circular, conforming to the bacterial pattern. A few groups of unicellular eukaryotes lack mitochondria: the microsporidians, metamonads, and archamoebae. On rRNA trees these groups appeared as the most primitive eukaryotes, suggesting they appeared before the origin of mitochondria, but this is now known to be an artifact of long branch attraction — they are apparently derived groups and retain genes or organelles derived from mitochondria (*e.g.* mitosomes and hydrogenosomes) (Henze and Martin, 2003). There are no primitively amitochondriate eukaryotes, and so the origin of mitochondria may have played a critical part in the development of eukaryotic cells.

The theory originally proposed by Lynn Margulis in the 1960s (Margulis, 1975) is now accepted – that mitochondria originated as endosymbiotic bacteria, taken up into proto-eukaryotic cells about 1.5 billion years ago, and provided energy generation in return for a safe environment. This prokaryotic past has left its traces in the many features of mtDNA which resemble those of modern bacteria:

- Circular, rather than linear, genome;
- Absence of histones;
- Discrete origins of replication, unlike nuclear chromosomes;
- No introns in genes, no dispersed repeats, and very little intergenic DNA;
- Polycistronic transcripts: transcription starts at only two promoters, one on each strand, and continues round the entire genome to produce RNA molecule from several genes;
- Different genetic code – perhaps the most striking evidence of the exogenous origin of mtDNA. In mammals, five codons have different specificities in mtDNA compared to the nuclear genome:

Codon	Nuclear code	mtDNA code
UGA	STOP	Trp
AGA, AGG	Arg	STOP
AUA, AUU	Ile	Met

Since the origin, mtDNA has lost most of its genes, and hence its autonomy. The 37 genes it now carries all play roles in either the oxidative phosphorylation pathway, or mitochondrial protein synthesis: 12S and 16S rRNAs and the 22 tRNAs required for mitochondrial protein synthesis plus 13 polypeptides of the mitochondrial energy generating process, oxidative phosphorylation (OXPHOS) (Figure 1). The other genes essential for mitochondrial function, including those encoding mtDNA polymerase, mtRNA polymerase and many structural and transport proteins, have been transferred to the nuclear genome. These nDNA-encoded mitochondrial proteins are translated on cytosolic ribosomes and selectively imported into the mitochondrion through various mitochondrial protein import systems. For example, certain proteins destined for the mitochondrial matrix are synthesized with an amino terminal, positively charged, amphoteric targeting peptide that is cleaved off once the protein enters the mitochondrial matrix (Wallace, 1999). The 13 mtDNA-encoded polypeptide genes are translated on mitochondrial ribosomes and all are structural subunits of OXPHOS enzyme complexes. These include 7 (ND1, 2, 3, 4L, 4, 5, 6) of the 46 polypeptides of complex I (NADH dehydrogenase), one (cytochrome b, cytb) of the 11 polypeptides of complex III (bc1 complex), 3 (COI, II, III) of the 13 polypeptides of complex IV (cytochrome c oxidase), and 2 (ATP 6 & 8) of the 16 proteins of complex V (ATP synthetase). The nDNA codes for all other mitochondrial proteins including all four subunits of complex II (succinate dehydrogenase), the mitochondrial DNA polymerase γ (POLG) subunits, the mitochondrial RNA polymerase components, the mitochondrial transcription factor (mtTFA), the mitochondrial ribosomal proteins and elongation factors, and the mitochondrial metabolic enzymes. The mtDNA is symmetrically transcribed from two promoters, one for the G-rich heavy (H) strand and the other for the C-rich light (L) strand. These H- and L-strand promoters (P_H and P_L) are contained in the approximately 1121-np control region (CR) that encompasses four mtTFA binding sites, the H-strand origin of replication (O_H), and three conserved sequence blocks (CSB I, II, and III). The L-strand origin of replication (O_L) is two thirds of the way around the mtDNA from O_H. Transcription is initiated at P_H or P_L and progresses around the mtDNA, generating a polycistronic message. The tRNAs, which punctuate the genes, are cleaved out and the mRNAs are then polyadenylated (Wallace, 2005). The non-coding control region (CR) of the mitochondrial genome is the displacement loop (D-loop) – triple stranded (extra 7S DNA), which contains a high degree of sequence variability between individuals. The highest degree of polymorphisms is concentrated within two hypervariable segments of the non-coding region, hypervariable segment I (HVS-I, 16024-16365 nt) and hypervariable segment II (HVS-II, 73-340 nt). The CR does not code for genes (Wilkinson-Herbots *et al.,* 1996).

The mtDNA gene repertoire has remained relatively constant since the formation of the fungal-animal lineage. This is because the mtDNA genetic code began to drift in the fungi such that the mtDNA genes can no longer be interpreted by the nuclear-cytosol system (Wallace, 1982). Consequently, when an mtDNA sequence is transferred to the nDNA, it remains as a nonfunctional pseudogene. As well as this ancient transfer of genes, segments of mtDNA have been inserted into nuclear genome at various points in our more recent evolutionary history. Indeed, a survey of the draft sequence of the nuclear genome (Bensasson *et al.,* 2001) indicates that it contains over 400 kb of mtDNA sequences – 25 times as much as the mtDNA molecule itself. Some of these nuclear mtDNA insertions (**numts**) are common to all primates, while others are more recent, some even being polymorphic in human populations. Numts are a potential source of problems in mtDNA studies in ancient samples, but also have their uses in the rooting of phylogenetic trees constructed using mtDNA (Mishmar *et al.,* 2004).

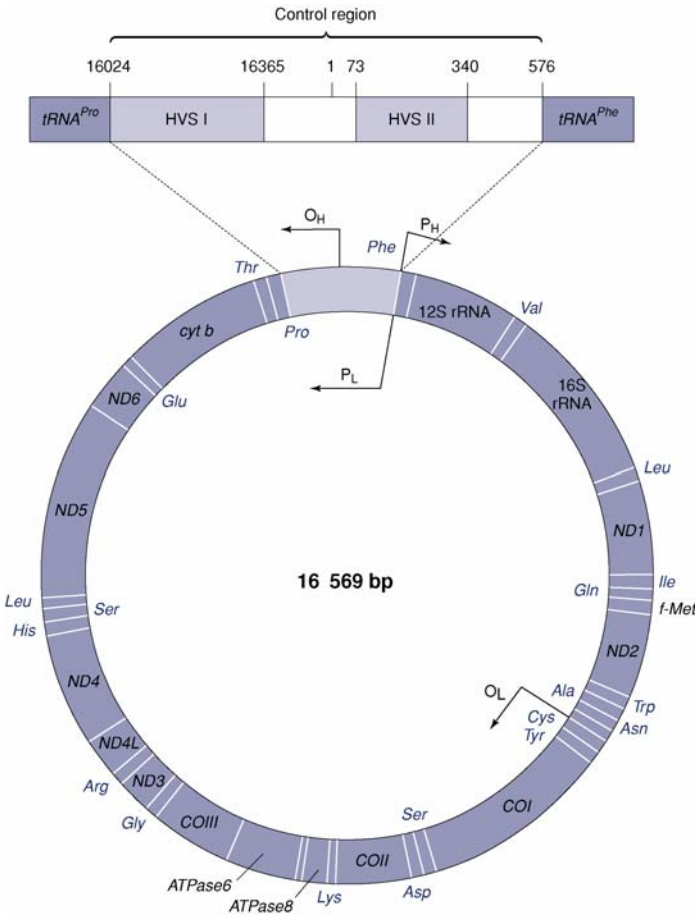


Figure 1. Human mitochondrial DNA (Jobling, Hurles and Tyler-Smith, 2004).
Human mtDNA is a circular double-stranded molecule with one strand (the heavy strand) relatively rich in G bases, and the other (the light strand) rich in C. tRNA genes are indicated by three-letter name of the corresponding amino acid; remaining genes are protein coding, except two rRNA genes. Origins of replication of the light (O_L) and heavy (O_H) strands, and promoters for transcription of these two strands (P_L, P_H) are shown. The control region contains two hypervariable segments (also known as hypervariable regions, HVRs) that are commonly assayed for variability (HVS-I and HVS-II). Genes that are transcribed from the H-strand and L-strand are shown outside or inside the circle. Arrows indicate the location of promoters for transcription and replication origins for both strands.

2.2. Special features of mtDNA

mtDNA, compared to eukaryotic nuclear genes and nuclear DNA in general, has a number of distinctive characteristics that makes it a useful marker for phylogenetic studies. There are maternal mode of inheritance, lack of recombination, homoplasmy and a relatively high mutation rate. A special aspect is the neutrality of its evolution – a problem that has been discussed extensively again very recently.

2.2.1. Maternal mode of inheritance and the lack of recombination

From the early of the 80-ties human population geneticists normally accept without question that human mtDNA are clonally inherited through the maternal line. The genetic evidence for this comes largely from pedigree analyses (*e.g.* Giles *et al.,* 1980). Inheritance of animal mtDNA is almost exclusively maternal, most likely because sperm-derived mitochondria are actively

eliminated from the ovum, either at or soon after fertilization (Manfredi *et al*., 1997). Maternal inheritance of mitochondrial DNA has long been regarded as a major paradox in developmental biology. While some confusion may still persist in popular science, research data clearly document that the paternal sperm-borne mitochondria of most mammalian species enter the ooplasm at fertilization and are specifically targeted for degradation by the resident ubiquitin system. Ubiquitin is a proteolytic chaperone that forms covalently linked polyubiquitin chains on the targeted proteinaceous substrates. The polyubiquitin tag redirects the substrate proteins to a 26-S proteasome, a multi-subunit proteolytic organelle. Thus, specific proteasomal inhibitors reversibly block sperm mitochondrial degradation in ooplasm. Prohibitin, the major protein of the inner mitochondrial membrane, appears to be ubiquitinated in the sperm mitochondria. (Sutkovsky *et al.,* 1999; 2004). Another mechanism of maternal mitochondrial inheritance in animals involves the selective elimination of sperm mitochondria by the elimination factor of the egg and the sperm mitochondria-specific factor. *In vitro* fertilization using sperm from isogenic mice incorporating heterospecific mtDNA showed that the elimination factor, which is probably an endonuclease, is selectively received by the tpis (tetratricopeptide repeat-containing protein involved in spermatogenesis) protein of the sperm mitochondrial outer membrane within the egg. It is then transported into the sperm mitochondria by Tom22 and Tom40 (translocator of mitochondrial outer membrane), where it destroys the sperm mtDNA, establishing the maternal inheritance of mtDNA (Hayashida *et al.,* 2005).

The number of mitochondria in a cell varies with cell type: those requiring a lot of energy, such as nerve and muscle cells, contain thousands of mitochondria, each containing 2 – 10 copies of mtDNA, while other cell types may contain only a few hundred. Oocytes contain around 100 000 mitochondria, each containing a single mtDNA molecule, while sperm contain only about 50 – 75 (Chen *et al.,* 1995). Clearly, even if fertilization involved a complete mixing of paternal and maternal mtDNA molecules, the contribution of the father to the zygote's pool of mtDNA would be relatively small. However, most evidence suggests that his contribution is actually zero, and that mtDNA is exclusively maternally inherited. Sperm mitochondria, required for motility, are localized in the midpiece, between the sperm head and tail. After fertilization the midpiece and its paternal mitochondria can be seen within zygote, and observed through several cell divisions. The apparent absence of paternal mtDNA inheritance suggests that there may be an active system to eliminate paternal mitochondria.

Maternal mode of inheritance and the lack of recombination (Olivo *et al.,* 1983; Merriwehter *et al.,* 1991; Eslon *et al.,* 2001; Piganeau and Eyre-Walker, 2004) allow to track individual genealogies and their evolution through the genetic history of human populations. There has been considerable debate about whether recombination occurs in mitochondrial DNA. Recombination has never been directly observed in human mtDNA but paternal inheritance has been recently observed (Schwartz and Vissing, 2002). Two lines of indirect evidence suggest that recombination might have occurred. The first line of evidence comes from the excess of homoplasies that are observed in phylogenetic trees, that is, identical mutation events occurring independently in different parts of a phylogeny (Eyre-Walker *et al.*, 1999). However, this excess of homoplasies could also be due to hypervariable sites and it thus remains unclear whether recombination or heterogeneity in the mutation rate is involved (McVean, 2001; Wiuf, 2001). The second line of evidence for recombination comes from the observation of a negative relationship between linkage disequilibrium (LD) and distance in some human mtDNA data sets (Awadalla *et al*., 1999). However, the analysis of other data sets has not corroborated this observation (Ingman *et al*., 2000; Jorde and Bamshad, 2000; Elson *et al*., 2001; Herrnstadt *et al*., 2002), and if the relationship is observed, it is only observed when LD is measured with r^2 as opposed to $|D'|$ (Jorde and Bamshad, 2000). This has led to an intense discussion about whether recombination occurs in human mitochondria (Jorde and Bamshad, 2000; Kivisild and Villems, 2000; Kumar *et al*., 2000; Parsons and Irwin, 2000; McVean, 2001; Wiuf, 2001; Innan and Nordborg, 2002). McVean (2001) suggested a method to get consistent result between the two statistics. This is to perform the analysis only on pairs of sites that are informative about

recombination. However, the choice of the informative sites requires a prior knowledge of the recombination rate in the data set analyzed, which makes the interpretation of the outcome of this test difficult. Further work on methods of detecting recombination from polymorphism data showed that the power for detecting recombination, that is, the probability of detecting recombination when there is recombination, by estimating the number of homoplasies (Posada and Crandall, 2001) or the relationship between LD and distance (Meunier and Eyre-Walker, 2001; Wiuf, 2001) is very low for small rates of recombination.
Due to the lack of recombination, mtDNA acts as a single locus. The effective population size of the mitochondrial genome is only one fourth that of autosomal loci. The influence of genetic drift thus makes mtDNA more sensitive to random fluctuations of allele frequencies than that for autosomal loci.

2.2.2. Mutations and mutation rate in human mtDNA

There are a number of possible reasons for the high rate of mutation of mtDNA compared to nuclear DNA:

- Because of its function in energy generation through oxidative phosphorylation, the mitochondrion contains a high concentration of mutagenic oxygen free radicals;
- mtDNA may have a higher turnover rate than nuclear DNA, requiring more replications per unit of time;
- because of the way it is replicated, mtDNA spends more time in the vulnerable single-stranded form than does nuclear DNA; indeed, the control region is sometimes called the "D-loop" region to reflect its unusual structure;
- mtDNA is not packaged with histones, which may make it more vulnerable to mutation.

Repair systems in mitochondria are less effective than those in the nucleus (reviewed by Shadel and Clayton, 1997; Bogenhagen, 1999). However, many elements of the mtDNA repair system have been shown to exist in animal mitochondria (Mason and Lightowlers, 2003).
Mutation in mtDNA warrants separate discussion because it has two unusual aspects – its rate with respect to nuclear DNA and the manner in which mutations pass from one generation to the next. Mutations occurring in mtDNA are largely base substitutions, and deletions are the causes of maternally inherited mitochondrial diseases, while many base substitutions appear to be neutral. Evolutionary genetic studies concentrate on these, but also consider other apparently neutral variants, such as change in length of a poly (C) tract in the control region, and the "9-bp deletion", a change in copy number of a 9-bp tandem repeat motif from two copies to one copy (rare mtDNAs even have three repeats). Early comparisons of human and other primate DNAs indicated that the base substitutional mutation rate in mtDNA is about 10 times higher than the average rate in nuclear DNA (Brown *et al.,* 1979). The high rate leads to a very large number of different sequences in human populations – one of the reasons for the popularity of mtDNA as an evolutionary tool. The observed mutation rate throughout the 16.5-kb molecule is non-uniform, with a relatively low rate in the coding region, and a relatively high rate in the coding region, which does not encode for proteins or RNAs. At some positions within the hypervariable segments (HVS-I and HVS-II), of the control region, mutation rate is so high that mutations can often be observed in pedigrees, and comparisons between the pedigree rate and the rate calculated by other means have caused controversy.
A good estimate of mtDNA mutation rate is important for human evolutionary genetic studies. The widely employed estimates are indirect, based on calibrating the amount of divergence or diversity accumulated since a well-dated event, such as a phylogenetic split between humans and a closely related primate, or a particular human settlement, such as that of New Guinea. Most interest has focused on the mutation rate of the control region, since this is the part of mtDNA most widely used in population studies. mtDNA mutation rates are often expressed as base substitutions per site per million years, rather than the "per base per generation" expression that has been used for nuclear DNA. A range of rates covering as order of magnitude has been

provided by these studies: 0.025 – 0.26 per site per million years (these figures include confidence intervals, and were compiled from the literature by Parsons *et al.,* 1997). This range assuming a 20-year generation time, equates to approximately 5 x 10^{-7} to 5 x 10^{-6} per base per generation. Some mutation rates focus on an even smaller region, the more variable of the two hypervariable segments – HVS-I. An average rate for a 275-bp section is given as one transition per 20 180 years (Richards *et al.,* 2000), which equates to 3.6 x 10^{-6} per base per generation, for a 20-year generation time. Outside the control region mutation rates have again been calculated by phylogenetic comparisons and assumptions about species divergence times: a rate from whole-mtDNA genome sequencing (Ingman *et al.,* 2000) is 3.4 x 10^{-7}.

Observations of mtDNA mutations, both neutral and disease-associated, indicate that the population of mitochondria passes through a bottleneck, at some point in oogenesis between the primordial germ cells and the primary oocyte. The result is that a mutant mtDNA representing a small minority of molecules in the soma of a mother can come to represent a range of proportions from zero to a large majority in her children. This phenomenon, known as cytoplasmic segregation, accounts for the wide variation in severity of mtDNA-associated diseases from one generation to the next, and also leads to interpretative problems when mtDNA mutation rates are considered. The term heteroplasmy is used to refer to a situation where more than one mtDNA type occurs in a cell, and homoplasmy when all mtDNAs are identical. Clearly, it is difficult to apply these terms rigorously, since in a large population of mtDNA molecules within a cell at any one time there may be several mutant varieties present at undetectable low concentrations; the degree of heteroplasmy may also be tissue-specific (Lightowlers *et al.,* 1997).

The substitution rate in mtDNA is strongly dependent on the region considered, and slow- and fast-evolving regions can be identified. (1) Nonsynonymous sites, the D-loop central domain, and tRNA and rRNA genes evolve much more slowly than synonymous sites and the two peripheral D-loop region domains. The synonymous rate is fairly uniform over the genome, whereas the rate of nonsynonymous sites depends on functional constraints and therefore differs considerably between genes. (2) The commonly accepted statement that mtDNA evolves more rapidly than nuclear DNA is valid only for some regions, thus it should be referred to specific mitochondrial components. In particular, nonsynonymous sites show comparable rates in mitochondrial and nuclear genes; synonymous sites and small rRNA evolve about 20 times more rapidly and tRNAs about 100 times more rapidly in mitochondria than in their nuclear counterpart. (3) A species-specific evolution is particularly evident in the D-loop region. As the divergence times of the organism pairs under consideration are known with sufficient accuracy, absolute nucleotide substitution rates are also provided (Pesole *et al.,* 1999). The considerable rate variation has also been observed between different D-loop nps (Hasegawa *et al.,* 1993; Macaulay *et al.,* 1997; Finnilä *et al.,* 2001). Transitions at nps 16093, 16129, 16189, 16311 and 16362 in HVS-I and 73, 146, 150, 152, 195 in HVS-II are considered as mutational "hotspots" and are often observed in different phylogenetic branches of mtDNA.

In phylogenetic tree-building, differences in mutation rates can be normalized by assigning different weights to the nps with known rate variation (Richards *et al.,* 1998). However, "hidden" and/or parallel mutations do not likely harm the outcome in any profound way, provided the level of resolution between the branches of an mtDNA phylogenetic tree is sufficient. The combined use of information both from fast evolving control region sequences and diagnostic coding region sites has justified itself in many mtDNA population genetic studies (*e.g.* Torroni *et al.,* 1996; Richards *et al.,* 1998, 2000; Macaulay *et al.,* 1999a; Kivisild *et al.,* 2000).

2.2.3. Natural selection, evolution of human mtDNA

Meiotic recombination is a consequence of sexual reproduction, and enhances the ability of populations to adapt to their environment through the combining of advantageous alleles at different loci. By contrast, non-recombining portions of the human genome (mtDNA and MSY)

are prone to the operation of “Muller’s ratchet”, the slow but inexorable accumulation of deleterious mutations. This process of generation may explain the low density of functional genes on the non-recombining portion of Y chromosome. While alleles at loci on different chromosomes are randomly segregated during meiosis, alleles at loci closely linked on the same chromosome are not, as recombination between them occurs infrequently. Linked loci share a common evolutionary heritage; selection operating on one locus will affect diversity at the other. An allele that rises to high frequency through positive selection at a linked locus is said to be “hitchhiking”. The reduction in diversity at loci linked to a recently fixed allele is dubbed a selective sweep. Conversely, negative selection at a locus also reduces diversity at linked loci, albeit at a slow rate, by a process known as background selection (Jobling, Hurles and Tyler-Smith, 2004).

A coherent theory of neutral evolution was first formalized by Kimura (1968). According to Kimura, when one compares the genomes of existing species, the vast majority of molecular differences are selectively “neutral”. That is, the molecular changes represented by these differences do not influence the fitness of the individual organism. As a result, the theory regards these genomic features as neither subject to, nor explicable by, natural selection. This view is based in part on the degenerate genetic code, in which sequences of three nucleotides (codons) may differ and yet encode the same amino acid (GCC and GCA both encode alanine, for example). Consequently, many potential single-nucleotide changes are in effect “silent” or synonymous substitution. Such changes are presumed to have little or no biological effect. However, it should be noted that the original theory was based on the consistency in rates of amino acid changes, and hypothesized that the majority of those changes too were neutral.

A second assertion or hypothesis of the neutral theory is that most evolutionary change is the result of genetic drift acting on neutral alleles. A new allele arises typically through the spontaneous mutation of a single nucleotide within the sequence of a gene. In single-celled organisms, such an event immediately contributes a new allele to the population, and this allele is subject to drift. In sexually reproducing, multicellular organisms, the nucleotide substitution must arise within one of the many sex cells that an individual carries. Then only if that sex cell participates in the genesis of an embryo and offspring does the mutation contribute a new allele to the population. Neutral substitutions create new neutral alleles. Through drift, these new alleles may become more common within the population. They may subsequently be lost, or in rare cases they may become "fixed"- meaning that their substitution becomes a 'permanent' feature of the population. When an allele carrying a new substitution becomes fixed, the effect is to add a new allele to the population. In this way, neutral substitutions tend to accumulate, and genomes tend to evolve. According to the mathematics of drift, when looking between divergent populations, most of the single-nucleotide differences can be assumed to have accumulated at the same rate as individuals with mutations are born. This latter rate, it has been argued, is predictable from the error rate of the enzymes that carry out DNA replication – enzymes that have been well studied and are highly conserved across all species.

Occurring deleterious mutations are removed by purifying selection; positive selection does not play any significant role. Theoretically, the rate of evolution solely depends on the mutation rate. The simplified model for explaining the present mtDNA variation is the following: mutations have accumulated sequentially along radiating female lineages and have reached to polymorphic frequencies only because of random genetic drift and its various manifestations, whereas the influence of (positive) natural selection has been negligible. According to the “near-to-neutral” theory of evolution there may also be, among the mutations of recent origin in the evolutionary time scale, slightly deleterious ones that are not yet removed by purifying selection (Hasegawa *et al.,* 1998; Nachman, 1998; Gerber *et al.,* 2001). The lack of recombination in human mtDNA makes it a subject of “Muller’s ratchet” – of a genetic mechanism that predicts an accumulation of slightly deleterious mutations. Many authors have tested the neutrality of mtDNA evolution by estimating the differences between the fixation of nonsynonymous and synonymous substitutions in the mtDNA of different species. They have found that there is an excess of

nonsynonymous mtDNA polymorphisms relative to fixed sequence change (*e.g.* Nachman, 1998; Excoffier and Yang, 1999). Studies addressing sequence variation in the mtDNA coding region have suggested that natural selection has significantly shaped the course of human mtDNA evolution. There was a significant excess of synonymous mutations in all genes coded by mtDNA, especially among those positions that defined the deeper branches of the tree (Cann *et al.,* 1984; Nachman *et al.,* 1996; Ingman and Gyllensten, 2001; Mishmar *et al.,* 2003; Moilanen *et al.,* 2003; Moilanen and Majamaa, 2003; Elson *et al.,* 2004; Ruiz-Pesini *et al.,* 2004). These studies have disagreed, however, upon whether the distribution of specific human mtDNA clades or haplogroups is due to an adaptation to different climates or if their distribution is a function of random genetic drift assisted by purifying selection that eliminates nonsynonymous changes. In an attempt to clarify this disagreement and to study the mode of natural selection in mtDNA variation in human populations, Kivisild *et al.,* (2006) have provided a phylogenetic analysis of a global sample of mtDNAs and investigate the position, chemical nature, and geographic distribution of recurrent and frequent mutations in the mtDNA coding region. Phylogenetic analysis of 277 human mitochondrial genomes, performed by Kivisild and colleagues (2006), has revealed a significant ($p<0.01$) excess of rRNA and nonsynonymous base substitutions among hotspots of recurrent mutation. Most hotspots involved transitions from guanine to adenine that, with thymine-to-cytosine transitions, illustrate the asymmetric bias in codon usage at synonymous sites on the heavy-strand DNA. The mitochondrion-encoded $tRNA^{Thr}$ varied significantly more than any other tRNA gene. Threonine and valine codons were involved in 259 of the 414 amino acid replacements observed. The ratio of nonsynonymous changes from and to threonine and valine differed significantly ($p= 0.003$) between populations with neutral and populations with significantly negative Tajima's D values, independent of their geographic location. In contrast to a recent suggestion that the excess of nonsilent mutations is characteristic of Arctic populations (Ruiz-Pesini *et al.,* 2004), implying their role in cold adaptation, demonstrated that the surplus of nonsynonymous mutations is a general feature of the young branches of the phylogenetic tree, affecting also those that are found only in Africa. Besides, they have introduced a new calibration method of the mutation rate of synonymous transitions to estimate the coalescent times of mtDNA haplogroups. As calibrated over the observed transversion rate at synonymous and rRNA positions between human and chimpanzee the average mutation rate over all human mtDNA genes yielded one synonymous substitution per 6,764 years (Kivisild *et al.,* 2006).

Despite the evidence of departures from neutrality and high levels of homoplasy at the interspecies level, the phylogenetic approach for analyzing mtDNA sequence data at the intraspecies level remains viable because the reconstruction of the basic branches is robust and the excess of nonsynonymous substitutions affects mainly the terminal branches of the tree. Thus, the genealogy of maternal lineages provides a link between observable sequence variation and evolutionary events that have shaped this diversity.

2.3. Phylogenetic trees and networks

The tree is an intuitively attractive method for displaying the relationships between many kinds of variant entities. Sometimes the tree itself describes the actual ancestral relationships of these entities; it encapsulates the mechanism by which diversity arose. Such is the case for trees of separate species or non-recombining haplotypes of unique markers – for example, human mtDNA and Y-chromosomal SNPs. Often, however, the tree is a convenient tool for the graphical display of diversity. In such cases, it may imply a model for how diversity arose, but does not itself represent the mechanism. For example, a tree of populations implies a model whereby populations split (fission) from common ancestors, and subsequently do not mix. This certainly does not represent the reality of population evolution. Different phylogenetic methods exist that can be used for reconstructing phylogenetic trees from interspecific molecular data (*e.g.* Fitch, 1977; Felsenstein, 1981; Saitou and Nei, 1987).

Reconstructing phylogenies from intraspecific data (such as human mitochondrial DNA variation) is often a challenging task because of large sample sizes; small genetic distances between individuals; homoplasies resulting from parallel mutation; and reversals of character changing. Median networks are commonly used when homoplasies resulting from parallel mutation or reversions are frequent. This type of data includes mtDNA control region sequences (where the base substitution mutation rate is high). Median networks are prone to producing hyperdimensional cubes or reticulations when the number of taxa becomes large, which quickly make the network unintelligible. A coherent set of rules has therefore been developed to "reduce" the network by removing some reticulations through elimination of the least likely links. For smaller datasets (less than 100 HVS-I sequences) the reduced median network, which consists of almost all possible maximum parsimony trees, can be used (Bandelt *et al.,* 1995).
An alternative method for constructing networks with limited levels of reticulation, known as median joining, has also been developed (Bandelt *et al.,* 1999). The algorithm used to construct these median-joining networks is based on the limited introduction of likely ancestral sequences/haplotypes into a minimum spanning network of the observed sequences. Again, these likely ancestral sequences are identified through the calculation of median haplotypes. The median joining algorithm has the advantage of being applicable to multi-state markers and is useful for large datasets, but is more unreliable for phylogenies with long branches. Consequently, the median joining algorithm is most often used for closely related, intraspecific haplotypes.
Thus reduced median and median joining networks represent alternative methods for obtaining intelligible networks within limited amount of reticulation. The former method generates all possible ancestral sequences and then eliminates the least likely, whereas the latter method introduces limited numbers of the most likely ancestral sequences into a phylogeny of the observed sequences.

- **Rooting of the phylogenetic tree**

Rooted trees contain one taxon that can be defined as having the most ancestral divergence compared to all other taxa. This taxon is also known as an outgroup. The root of the tree lies between the outgroup an all other taxa. This property orientates rooted trees with respect to evolutionary time, meaning that evolutionary changes on the tree have a defined direction of change, from ancestral to derived. An unrooted tree can be rooted either by assuming that the root falls midway along the longest branch on the tree (mid-point rooting), or by incorporating a taxon known to be an outgroup to all other taxa, seeing where it joins the unrooted tree. The outgroup that already earlier in evolution has separated from the most recent common ancestor (MRCA) of the studied group will be chosen. In human mtDNA analysis, the corresponding sequences of chimpanzee (Vigilant *et al.,* 1991; Ingman *et al.,* 2000; Maca-Mayer *et al.,* 2001), Neanderthal (Krings *et al.,* 1997; Ovchinnikov *et al.,* 2000), as well as of the recent insertion of D-loop segment into nuclear genome (Watson *et al.,* 1997), have been used as an outgroup. If the outgroup can not be specified, the mid-point rooting will be employed. Mid-point rooting was used in the first studies of human mtDNA (Cann *et al.,* 1987), when the data of suitable outgroups was lacking.

2.4. Calibration of the mtDNA molecular clock

The molecular clock hypothesis holds that for any given DNA sequence, the rate of evolution is approximately constant over all evolutionary lineages. This regularity of molecular evolution would stand in direct contrast to the non-uniform change of morphological evolution. In principle, this hypothesis could be tested using known calibration points of dated lineage divergences. This requires accurate, independent dating of the lineages being investigated by other disciplines, most notably palaeontology. Mutation rate calibration can either be performed

directly on individual meioses, or indirectly through the observation of a certain amount of divergence across a known time span. The latter approach often uses divergence between species whose divergence is well dated in the fossil records.
One of the approaches is to estimate the mean rate of mtDNA divergence in different geographical regions for which approximate time of colonization is available from historical sources. The extent of differentiation within clusters specific to New Guinea, Australia and the Americas has been used and the divergence time (twice the substitution rate) estimate between 2-4% per million years has been calculated for whole human mtDNA molecule (Wilson *et al.,* 1985; Cann *et al.,* 1987; Torroni *et al.,* 1994a). The same value fro transitions of D-loop HVS-I region (between nps16090 to 16365) was found to be 36% per million years (Forster *et al.,* 1996).
The other opportunity is to use the outgroup method, which compares the average amount of sequence variation between two species considering their distance from the MRCA. The time of their split is taken from paleontological evidences.
Recently, Kivisild *et al.,* (2006) have performed calibration using nuclear inserts of mtDNA and the chimpanzee consensus sequence as outgroups appeared between haplogroup L0 and the rest of the phylogeny. Extensive interspecies homoplasy and mutational saturation was highlighted by the fact that for more than one-third (417/1292) of the variable sites, regardless of their phylogenetic position on the tree, the derived allele among humans corresponded to the chimpanzee allele. In agreement with noncoding region information (Aquadro and Greenberg, 1983), a high ratio (21.5 on average, 34.8 in synonymous positions) of transitions to transversions was observed in the coding region (577–16023). Interspecies calibration of the molecular clock over the complete mtDNA sequence (Ingman *et al.,* 2000; Mishmar *et al.,* 2003) is problematic because of saturation of transitions at silent positions and the effect of selection on the fixation rate of amino acid replacement mutations (Ho *et al.,* 2005). Assuming 6 million years for the human–chimpanzee species split (Goodman *et al.,* 1998) and 6.5 million years for the most recent common ancestor of their mtDNA lineages (Mishmar *et al.,* 2003), Kivisild *et al.,* (2006) estimated the average transversion rate at synonymous and rRNA positions as $2.1x10^{-9}$ and $4.1x10^{-10}$/ year/position, respectively. Using the observed relative rates of different substitution types in humans, the average transition rate at 4212 synonymous positions is $3.5x10^{-8}$ (SD $0.1x10^{-8}$)/year/position. Over all genes in mtDNA this would be equivalent to accumulation of one synonymous transition/6764 (SD 140) years on average. The coalescent date of the human mitochondrial DNA tree using this rate is 160,000 (SD 22,000) years. This coalescent date is broadly consistent with the dates of the *Homo sapiens* fossils recognized so far from Ethiopia (Clark *et al.,* 2003; White *et al.,* 2003; McDougall *et al.,* 2005). The TMRCA of all the Eurasian, American, Australian, Papua New Guinean, and African lineages in clade L3 dates to 65,000 ± 8000 years while the average coalescent time of the three basic non-African founding haplogroups M, N, and R is 45,000 years (Kivisild *et al.,* 2006). These estimates, bracketing the time period for the recent out-of-Africa migration (Stringer and Andrews, 1988), are younger than those based on calibrations involving all coding region sites (Ingman *et al.,* 2000; Mishmar *et al.,* 2003) but are still in agreement with the earliest archaeological signs of anatomically modern humans outside Africa (Mellars, 2004). The differences between the date estimates of previous studies are most likely due to the overrepresentation of possibly slightly deleterious nonsynonymous mutations in the younger branches of the tree (Elson *et al.,* 2004) that introduces a bias to the coalescent approach if all the sites of the coding region are used.

2.5. Early studies of phylogeographic diversity of human mtDNA

The first application of mtDNA to elucidate the origin of modern humans took place in the late 1970s and early 1980s, when Brown *et al.,* (1980) discovered the restriction fragment pattern of mtDNAs among individuals from diverse ethnic and geographical origin differed substantially. On the basis of an estimated rate for base substitution of 1%/106 year (Brown *et al.,* 1979), man

could have speciated or passed through a severe population constriction as recently as 180,000 years ago. The complete sequence of mitochondrial genome was published in 1981 (Anderson *et al.,* 1981) and soon after that many studies of different population were undertaken. While Brown *et al.,* (1980) had used 18 restriction endonucleases and 21 samples, Denaro *et al*. (1981) treated a large number of samples (235) from five ethnic groups with only a single restriction enzyme. They found that the presence of *Hpa*I restriction site that corresponds to the transition at np 3954 separates most of the Africans from the Eurasian individuals. Notably, the Africans had a derived sate of this np when compared to other primates, whereas the rest of studied population shared an ancestral state. Based on these findings was proposed, Asia is possible genetically central to the mtDNA radiations that are thought to have given rise to the human ethnic groups. This study was supported by the investigation of Blanc *et al.,* (1983) who discovered the frequent presence of ancestral state of one *Hinc*II polymorphism (corresponding to transition in np 12406) among Asians and was upheld by studies of Nepalese (Brega *et al.,* 1986) and Chinese (Yu *et al.,* 1988) populations. Phylogenetic analysis of the Japanese population concealed a considerably high degree of mtDNA diversity (Horai *et al.,* 1984). Cann *et al.,* (1982) demonstrated that mtDNA diversity among aboriginal Australians is as diverse as in any other populations tested in the Old World. Johnson *et al.,* (1983) showed that the mtDNA diversity is greatest in Africa, and that all mtDNA variants present toady can be seen as deriving from a single phylogenetic tree. Although they argued that the highest diversity in Africa can be caused by the longer age of African variants as well as by different mutation rates in different mtDNA lineages, the mid-point root of their mtDNA tree indicated the African origin. Later, Excoffier and Langaney (1989) suggested that the high diversity of African could be explained equally well by assuming that selection has played a significant role in creating the present mtDNA variation. The possibility to trace the origin and to study the demographic history of human populations by the use of mtDNA as a population specific marker became highly visible in 1987 (Cann *et al.,* 1987). They used high-resolution restriction analysis of 147 mtDNAs from five geographic populations and concluded that the root of human mtDNA phylogenetic tree is in Africa as well as all the populations examined except the African populations have multiple origins, implying that each area was colonised repeatedly. Compared to earlier studies (cited above) that allowed the screening of 2-4% of the total mtDNA sequence variation, in this analysis a ten time higher resolution level was obtained.

2.5.1. Hypotheses to explain the origin of modern humans

- Multiregional model of human evolution

The results of early mtDNA studies led to construction of a starlike mtDNA tree with a central node shared by a high number of individuals from all over the world (Excoffier and Langaney, 1989). Other lineages, some of which population specific, radiated form that central haplotype. The root of this tree was disputed and it was sometimes interpreted as support for the multiregional model of the origin of modern humans (Templeton, 1992). This model is based mostly on archaeological evidence and interpretation of morphological markers of fossil records. The multiregional model was first proposed by F. Weidenreich (1943), but was futher developed and promoted by Milford Wolpoff and colleagues (Wolpoff *et al.,* 1984; Wolpoff and Caspari, 1997). The following model proposes that the transition from *H. erectus* to *H. sapiens* took place in a number of places in the Old World, with the diverse modern human characteristics arising at different times in different places. These scientists explained the apparent regional continuity by claiming that *Homo erectus*, Neanderthals, *Homo sapiens* and other humans were a single species. This species arose in Africa two million years ago as *Homo erectus* and then spread out over the world, developing adaptations to regional conditions. For periods of time some populations became isolated, developing in a different direction. But through a complicated process involving continuous interbreeding, replacement, genetic drift and other vehicles of

evolution, adaptations that were an advantage anywhere on earth would spread, keeping the development of the species in the same overall direction, while maintaining adaptations to regional factors. Eventually, the more unusual local varieties of the species would have disappeared in favour of modern humans while retaining some regional adaptations, but also with many common features.

- "Out of Africa": recent African origin and anatomically modern humans

In contrast to multiregional model, the "recent out of Africa" model (known also as *Garden of Eden* model) proposes that the *Homo erectus* to *Homo sapiens* transition took place recently (<200 KYA) in Africa, and that AMHs replaced the archaic hominids already present on other continent. This theory was proposed in its extreme form – assuming total replacement of all archaic ancestors – by archaeologists Stringer and Andrews (1988) at about the same time when the genetic evidences of recent African origin of human mtDNA appeared (Cann *et al.,* 1987). The most important results of this study carried out by Wilson's group were: 1) the midpoint-rooted maximum parsimony tree obtained from the results of the high resolution RFLP analysis was not starlike – it revealed a deep split between the two main branches, one of which consisted only of African lineages and the other encompassed both African and non-Africans; 2) the mtDNA variation was highest in Africa. These findings were interpreted as evidence of the African origin of mtDNA in extant populations – all mtDNAs stem from one woman ("*mitochondrial Eve"*) who lived *ca* 200,000 years ago. The original paper by Cann et al. (1987) was criticized by mainly on five grounds: (1) RFLP analysis is not suited to estimate mutation rates, which is essential in timing evolutionary events; (2) of the 147 persons sampled, only two were from sub-Saharan Africa. The other 18 'Africans' in the study were Afro-Americans; (3) the method to generate the tree was not guaranteed to find the most parsimonious tree; (4) the method used to root the tree placed it at the midpoint of the longest branch (midpoint rooting). This could lead to a wrong position of the root if for example the rate of evolution is higher in Africa; (5) weak statistical analysis (Darlu and Tassy, 1987; Saitou and Omoto, 1987; Excoffier and Langaney, 1989; Kruger and Vogel, 1989; Maddison, 1991). Bearing in mind the criticism of their first analysis, Wilson's group (Vigilant *et al.,* 1991) analyzed 189 sequences of two hypervariable segments of mtDNA control region, including those from 121 native Africans. The branches of the maximum parsimony tree, rooted by the more powerful outgroup-method, which used chimpanzee sequence, started again exclusively from Africa. The calibration of this tree with the mutation rate calculated from comparisons of average amount of sequence differences between human and the outgroup, resulted in approximately the same age of "mitochondrial Eve" that was obtained in the previous publication (Cann *et al.,* 1987). However, criticisms of those studies stayed unmoved because equally parsimonious trees from the same dataset showing other results were obtained (Hedges *et al.,* 1992; Templeton, 1992). The conclusion that could be drawn from those early papers was that the available sequence data and the genealogical resolution level were insufficient to solve statistically the place of origin of human mtDNA.
Nowadays, the recent out-of-Africa model has found predominant support from the mtDNA analysis of large datasets analyzed with different methods (*e.g.* Chen *et al.,* 1995; 2000; Horai *et al.,* 1995; Watson *et al.,* 1997; Ingman *et al.,* 2000). A more recent mtDNA phylogeny bases on the complete mtDNA sequences of 53 individuals of diverse geographical origins, rooted by comparison to a chimpanzee sequence, was determined by Ingman *et al.,* (2000). All sequences were different and 675 variable positions were found, 516 of which were outside the hypervariable region. Despite the elevated mutation rate of mtDNA compared with nuclear sequences, a robust phylogeny could be obtained from the complete sequence excluding the rapidly mutating control region. It has some striking features:

- Complete separation of African and non-African lineages;
- The first three branches lead exclusively to African lineages, while the fourth branch contains both African and non-African lineages;

- Deep branches within African lineages; starlike structure within non-African lineages;
- TMRCA for the entire phylogeny: 172 ± 50 KY;
- TMRCA for the branch containing African plus non-African lineages: 52 ± 28 KY;
- Expansion time for non-African lineages estimated at 1925 generation or 38 500 years at 20 years/generation.

Based on the deepest splits of the phylogeny constructed from 277mtDNA complete coding region sequences were sustained by African mtDNAs, which belonged to previously defined haplogroups L0–L5, and the coalescent date of the human mitochondrial DNA tree was calculated to be 160,000 (SD 22,000) years (Kivisild *et al.,* 2006). However, the out-of-Africa model is certainly over-simplified, as it does not account for continental subdivisions, environmental changes affecting the demographic history of populations, and lacks mechanism to explain the present diversity of AMH.

Intermediate models are obviously possible, involving a recent origin of many human characteristics in Africa, but also interbreeding with archaic populations outside Africa. Relethford (2001) has suggested that it is useful to distinguish between the mode of transition (multiregional evolution within species, or a speciation event followed by replacement) and the location and timing of the transition (no single time and place, or recent in Africa). As a result, he classifies models into three groups: regional coalescence, primary African origin, and African replacement. We can also come across the terms "strong garden of Eden hypothesis" and "weak garden of Eden hypothesis" (Harpending *et al.,* 1993). These refer to two demographic variants of the out of Africa model. According to the "strong" hypothesis, modern humans appeared as an *H. erectus* subpopulation, perhaps, a new species, and spread continuously over much of the Old World. According to the "weak" hypothesis, modern humans again appeared in one subpopulation, but spread was slow, taking tens of thousands of years and producing separate daughter populations that later re-expanded. Various alternative modes and pathways of dispersal of AMHs, following the northern (over Sinai) and/or southern (over southern Arabia) routes have been put forward, based on paleoanthropoligical (Lahr and Foley, 1998; Stringer, 2003) and genetic data (Cavalli-Sforza *et al.,* 1994; Hammer *et al.,* 1997; 1998; Jin *et al.,* 1999; Kivisild *et al.,* 1999a; 2000; 2003a,b; Quintana-Murci *et al.,* 1999; Underhill *et al.,* 2000; Cann *et al.,* 2001; Templeton, 2002) (Figure 2). In conclusion, AMHs originated from a recent evolutionary event, whereas modern human diversity is the result of multiple evolutionary events brought about by multiple geographic dispersals.

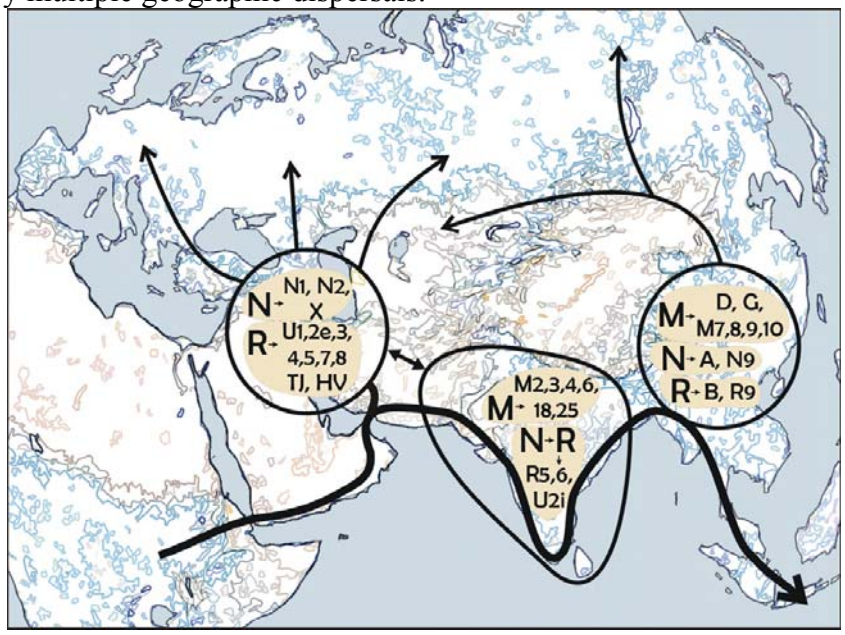


Figure 2. Two proposed routes out of Africa – the Northern and the Southern route (Metspalu *et al.,* 2004). The bold black arrow indicates the possible "coastal" route of colonization of Eurasia by anatomically modern humans (ca. 60,000 – 80,000 YBP.). This "Southern Coastal Route" is suggested by

the phylogeography of mtDNA haplogroup M, the virtual absence of which in the Near East and Southwest Asia undermines the likelihood of the initial colonization of Eurasia taking a route north around the Red Sea. Therefore, the initial split between West and East Eurasian mtDNAs is postulated between the Indus Valley and Southwest Asia. Spheres depict expansion zones where, after the initial (coastal) peopling of the continent, local branches of the mtDNA tree (haplogroups given in the spheres) arose (ca. 40,000 – 60,000 YBP), and from where they where further carried into the interior of the continent (thinner black arrows). Admixture between the expansion zones has been surprisingly limited ever since.

2.6. Global variation. Nomenclature of mtDNA haplogroups

The first studies of mtDNA variation in human populations used either the RFLP analysis or sequencing of the hypervariable segments (mostly HVS-I) of the control region. A more refined picture about the spread of different mtDNA variants started to emerge during early 1990s with the application of the high-resolution restriction fragment analysis to study the mtDNA variation from one continent at a time (*e.g.* Ballinger *et al.,* 1992; Torroni *et al.,* 1992; 1993a;1993b; 1994a; 1994b; 1994c; Chen *et al.,* 1995). This series of investigations established "the backbone" of the human mtDNA phylogenetic tree that has largely been in used since then. Meanwhile, an independent classification was proposed, based on the phylogenetic analysis of the variation of HVS-I part of mtDNA (Richards *et al.,* 1996). Thereafter, a new series of studies allowed the combination of the control region and coding region (RFLP) data (Torroni *et al.,* 1996; Richards *et al.,* 1998; Macaulay *et al.,* 1999b; Shurr *et al.,* 1999). As a result, it was established that all mtDNA variants, or *haplotypes*, with characteristic polymorphism can be divided into a number of monophyletic clades, or *haplogroups* (hgs), defined by a single or several mtDNA coding region polymorphisms (often defined by a gain or a loss of a restriction site), associated with polymorphisms (if any) of the mtDNA control region. The main haplogroups are denoted by capital letters and consist of different sub-haplgroups defined by characteristic polymorphisms (Figure 3). Most of the haplogroups exhibit restricted geographical spread (table 1), thus distinguishing populations from different continents/geographical regions. The refinement of the mtDNA tree and the nomenclature of mtDNA is an ongoing process. In particular, data emerging from complete mtDNA sequences (Ingman *et al.,* 2000; Finnilä *et al.,* 2001; Finnilä and Majamaa, 2001; Maca-Meyer *et al.,* 2001; Herrnstadt *et al.,* 2002; Kivisild *et al.,* 2002; Yao *et al.,* 2002a; Kong *et al.,* 2003b; Reidla *et al.,* 2003) offer a better phylogenetic resolution of different variants of maternal lineages.

The likely root of human mtDNA tree is between haplogroups L0 and L1 dividing the phylogenetic tree into two basic clades: L0 and the rest. The oldest lineage cluster L was initially defined by the presence of the *Hpa*I restriction site at np 3592 (Chen *et al.,* 1995). This study was based on a phylogenetic analysis where Asian and European mtDNA sequences were used as the outgroups for the African tree. Therefore, L is not a "real" clade and includes several distinctive paraphyletic clusters of African mtDNA lineages. Using chimpanzee mtDNA sequence as an outgroup for human mtDNA, seven major African clades within L can now be recognized – L0-L6 (Bandelt *et al.,* 1995; Chen *et al.,* 1995, 2000; Graven *et al.,* 1995; Watson *et al.,* 1997; Alves-Silva *et al*. 2000; Torroni *et al.,* 2001b; Salas *et al*. 2002, 2004; Kivisild *et al.,* 2004, 2006, Beleza *et al.,* 2005; Coia *et al.,* 2005; Batini et al. 2006; Gonder *et al.,* 2006; Olivieri *et al.,* 2006). Each of these clades can be further divided into several subclusters (Figure 3). Different estimates of the age of MRCA for African lineages have given very similar results with coalescence values of 100,000-170,000 YBP (Chen *et al.,* 1995, 2000; Graven *et al.,* 1995; Watson *et al.,* 1997). The MRCA of the oldest clade containing both African and non-African individuals lies in hg L3 and dates to *ca* 50,000-80,000 YBP (Watson *et al.,* 1997; Ingman *et al.,* 2000).

It seems that only L3 radiated out of Africa, in the form of haplogroups M and N, about 60,000 YBP, giving rise to the extant Eurasian variation (Quintana-Murci *et al.,* 1999; Wallace *et al.,*

Region	Population	N	H	Pre-V+V	Pre-HV+HV	U	U*	U1	U2	U3	U4	U5	U6	U7	U8	K	T	J	R*	R9	B	N*	W	N1	X	A	N9	M*	Mi	C	D	G	Z	L0	L1	L2	L3
Europe	Finns[1]	403	40	5		31						27			1	3	2	4					9	3	1								2				
	Norwegians[2]	74	38	5		23	1			1		16			1	4	12	11					3	5	1												
	Russians[3]	201	42	5	2	21		1	1	1	3	11				4	11	8					2	2	3												
	Bosnians[4]	144	48	6	2	24		1		1	6	12				4	5	7					1	3	1			1									
	French[5]	320	48	4	2	23	2		2		1	8	1		1	8	11	6					2	3	1												
Near East	Turks[6]	387	25		6	26	1	4	1	5	1	5		2	1	6	11	11	2			1	4	5	5	1	1	1		1	1	1	1				
Siberia	Yakuts[7]	191	3		1	2					1	1					1	1		6	1		2			2	2	3		44	30	4					
	Altaians[8]	110	6			16			5	2	5	4					1	1		9	4			3	3		5	7		19	15	2	5				
Asia	Han Chinese[9]	263																	1	16	16					7	6	23		3	22	3	2				
	Thailand[10]	552				4			4										5	45	13					3	1	7			17	5					
	India[11]	1205	3		2	15		1	10		1	1		2			2	1	15	3			1					40	19		1						
America	Native Americans[12]	300																			25			2	2	33				20							
Africa	Northwest Africans[13]	268	18	2		16				1		4	8			3	4	3										1							13	25	17
	Mozambique[14]	416																															25	12	36	27
	Guinea-Bissau[15]	372				5						3	2															1					5	16	43	30
	Kung!/Khwe[16]	93																															4	70	8	18

Table 1. Frequencies (%) of major mtDNA haplogroups in different world populations

Note: Asterisk (*) indicates unspecified derivates of a particular haplogroup. Frequencies of hg U sub-hgs are given separately. Data are from: [1]Meinilä *et al.* 2001; [2]Passarino *et al.* 2002; [3]Malyarchuk *et al.* 2002; [4]Malyarchuk *et al.* 2003; [5]Cali *et al.* 2001; Dubut *et al.* 2004; [6]Tambets *et al.* 2000; [7]Fedorova *et al.* 2003; [8]Derenko *et al.* 2003; [9]Yao *et al.* 2002b; [10]Fucharoen *et al.* 2001; Oota *et al.* 2001; Yao *et al.* 2002a; [11]Bamshad *et al.* 2001; Kivisild *et al.* 2003; [12]Torroni *et al.* 1993b; [13]Rando *et al.* 1998; [14]Pereira *et al.* 2001; Salas *et al.* 2002; [15]Rosa *et al.* 2004; [16]Vigilant *et al.* 1991; Chen *et al.* 2000. Mi encompasses haplogroup M sub-clades spread mostly only in India.

1999). Most western Eurasian populations are characterized by clades within hg N (Torroni *et al.,* 1996; Macaulay *et al.,* 1999a; Richards *et al.,* 2000; Finnilä *et al.,* 2001; Herrnstadt *et al.,* 2002; Palanichamy *et al.,* 2004), whereas N and M contributed almost equally to the current eastern Eurasian mtDNA pool (Stoneking *et al.,* 1990; Ballinger *et al.,* 1992; Torroni *et al.,* 1993a,b; Horai *et al.,* 1995; Kolman and Bermingham, 1997; Comas *et al.,* 1998; Starikovskaya *et al.,* 1998; Derbeneva *et al.,* 2002b; Kivisild *et al.,* 2002; Shurr and Wallace, 2002;Yao *et al.,* 2002a,b).

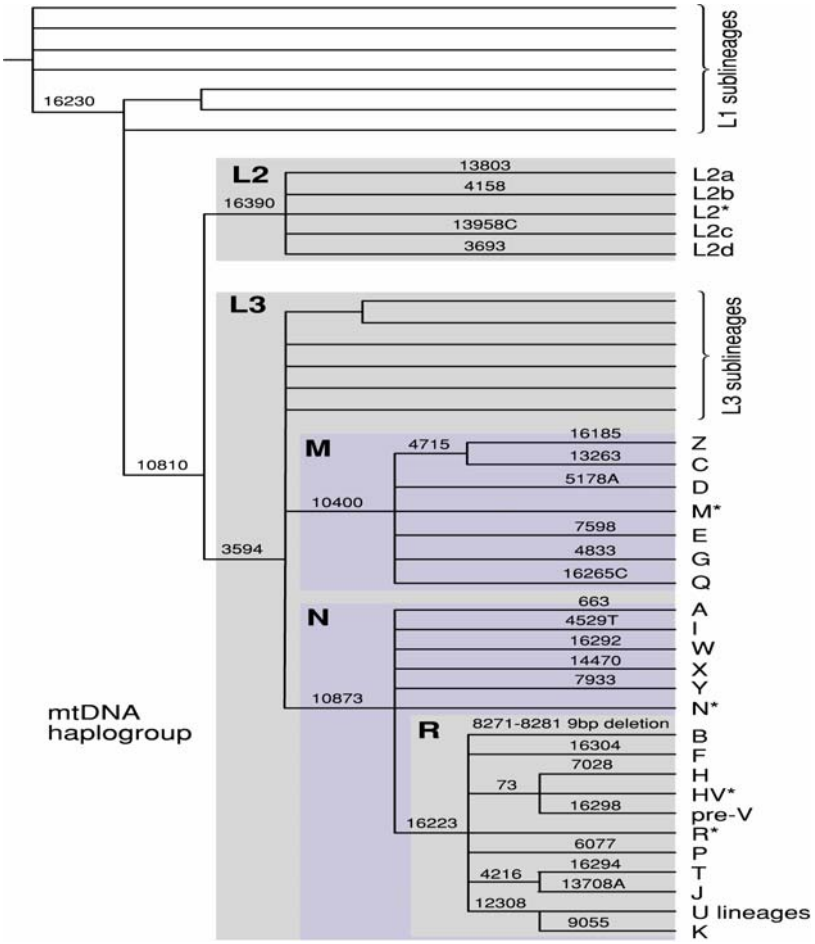


Figure 3. A high level phylogeny of human mitochondrial DNA (Jobling, Hurles and Tyler-Smith, 2004). Clades are labeled with position within the genome of one of the clade-defining mutations. All mutations are transitions with the exception of few transversions that are indicated by the nucleotide position being followed by the derived base. The topology of L1 and L3 sublineages (taken from Salas *et al.,* 2002) illustrates that L1 is not monophyletic. Data are from Macaulay *et al.,* 1999a; Schurr *et al.,* 1999; Chen *et al.,* 2000; Ingman *et al.,* 2000; Herrnstadt *et al.,* 2002; Kivisild *et al.,* 2002; Salas *et al.,* 2002; Mishmar *et al.,* 2003.

Because of its great time depth and virtual absence in western Eurasians, it is not excluded that hg M was brought to Asia from East Africa, along the southern route, by the earliest migration wave of AMHs (Kivisild *et al.,* 1999a; 2000; Quintana-Murci *et al.,* 1999). Distinct M variants make up about 20% of the mtDNAs in Ethiopia, leading Quintana-Murci and colleagues (Quintana-Murci *et al.,* 1999) to identify Eastern Africa as the source of migration out of Africa involving the ancestors of the current Indians and other Asian populations. It is possible to estimate dates for the nodes in the mtDNA phylogenetic tree. Quintana-Murci *et al.,* (1999) use the ρ method to estimate coalescent times of 36 ± 11 or 48 ± 15 KYA for the Ethiopian haplogroup M and 53 ± 7 or 56 ± 7 KYA for Indian haplogroup M chromosomes. They judge that these ancestors must themselves have diverged for several thousand years, so that the common ancestor of all M mtDNAs would have lived about 60 YBP. Recently, the same scenario was proposed for the spread of hg N (Kivisild *et al.,* 1999a; 2003a, b). The lack of L3 lineages other than M and N among non-Africans (Ingman *et al.,* 2000; Herrnstadt *et al.,* 2002;

Kivisild *et al.,* 2002) and, particularly, in South Africa, suggests that the earliest migration(s) of modern humans already carried these two mtDNA ancestors (Kivisild *et al.,* 2003b). This scenario is also consistent with the fact that the founder ages of M and N have been shown to be very similar – 54,200 ± 11,400 and 53,400 ± 11,700 YBP, respectively (Forster *et al.,* 2001). Also, it has been proposed by Kivisild *et al.,* (2003b) that, considering the mtDNA variation in South Asia, the N branch had relatively early given rise to its large daughter clade R, which later, among eastern Eurasians, differentiated into clusters B and R9 and gave rise to hgs HV, TJ, and U (figure 4) among western Eurasians.

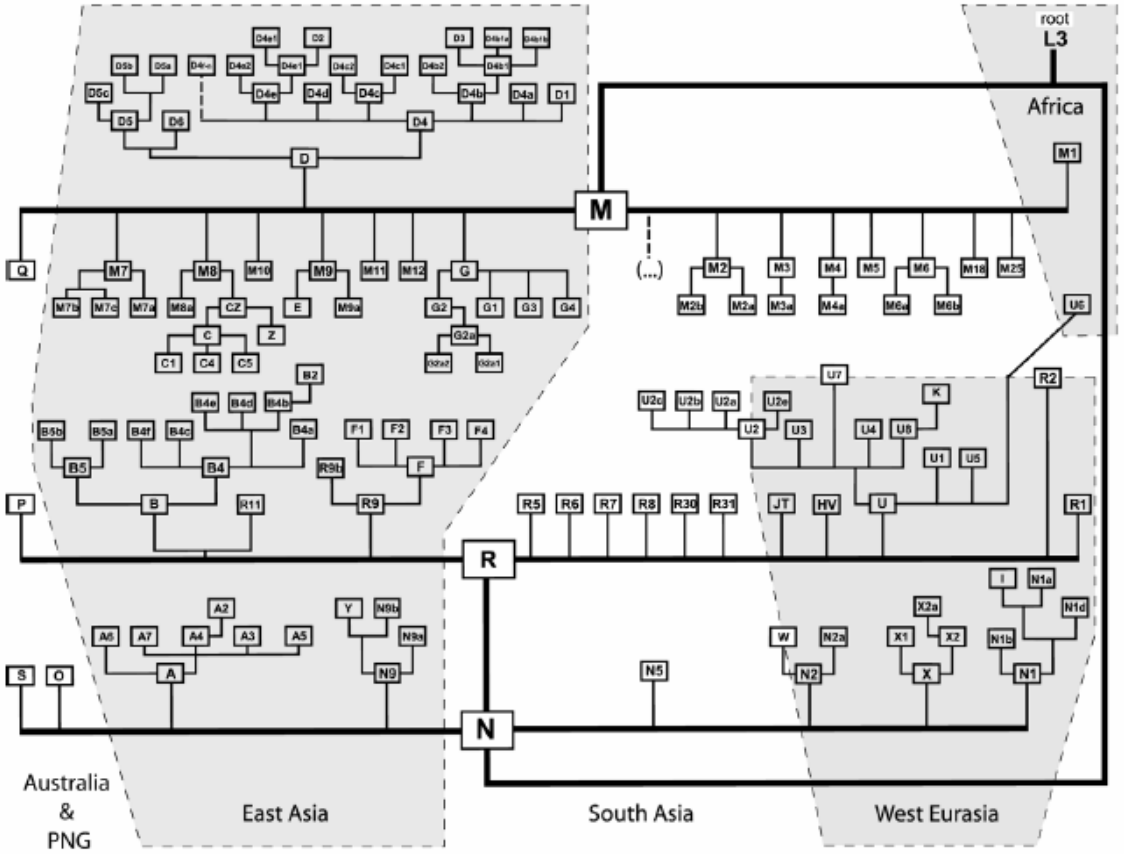


Figure 4. Schematic tree of human mtDNA haplogroups spread in eastern and western Eurasian populations (for eastern Eurasian based on Quintana-Murci *et al.,* 1999; Kivisild *et al.,* 2002; Yao *et al*. 2002b; Kong *et al.,* 2003b; Metspalu *et al.,* 2004; for western Eurasian based on Macaulay *et al.,* 1999a; Richards *et al.,* 2000; Finnilä *et al.,* 2001; Herrnstadt *et al.,* 2002; Reidla *et al.,* 2003; Quintana-Murci *et al.,* 2004; Qiuntans *et al.,* 2004). See www.mitomap.org (phylogenetic tree) for defining HVS-I and coding region nps. PNG – Papua New Guinea.

2.6.1. mtDNA variation in Asian populations

Complete and partial mtDNA coding region sequences have been used to determine the fine-structure of the mtDNA lineages present in Asia (Kivisild *et al.,* 2002; Yao *et al.,* 2002a,b; Kong *et al.,* 2003a; Metspalu *et al.,* 2004; Palanichamy *et al.,* 2004; Quintana-Murci *et al.,* 2004). The recent analysis of complete mtDNA sequences from 672 Japanese individuals has provided a significant refinement of the East Asian mtDNA phylogeny (Tanaka *et al.,* 2004). The initial southern (coastal) route of the out-of-Africa event had taken place by around 60-65 KYA (Maca-Meyer *et al.,* 2001; Mishmar *et al.,* 2003; Palanichamy *et al.,* 2004). During this opening stage, the earliest branches of hgs M and N were rapidly segregated in to West Asian (WA: *e.g.* JT, U), South Asian (SA: *e.g.* M2, N5, R5, U2a, U2b, U2c), East Asian (EA: *e.g.* D, M7, M8, N9, R9, B) and further into Australia-specific (Sahul: S, O, P, Q) variants.

Combining these and other published data, Figure 4 summarizes the Asian mtDNA tree topology. The macrohaplgroups M and N effectively cover the whole mtDNA pool in Asia. Macrohaplogroup M is slightly more frequent than N in Siberia, northern China, Japan and South Asia, while in Southeast Asia it is the other way around. M is practically absent from Southwest Asia where subhaplogroups branching from N (including R) dominate the mtDNA landscape. The N and R sub-branches in West and East Eurasia do not overlap, thus forming two distinct mtDNA domains. With approximately similar shares, these two make up most of the mtDNA pool of Central Asia (Comas *et al.,* 2004). Almost all eastern Eurasian R lineages belong to the two major hgs B and R9.

Sub-clades of R that encompass the majority of mtDNA variants spread in western Eurasia (HV, TJ, and U) are or absent in most eastern Eurasian populations. Eastern and western Eurasian-specific mtDNA packages meet in Central Asia, which is a contact zone between those regionally differentiated groups. There, the contribution of eastern and western Eurasian mtDNAs to the total mtDNA pool is more or less equal (Comas *et al.,* 2004; Quintana-Murci *et al.,* 2004). The contribution of western Eurasian components (U4, H, JT) account one third also in some western and southern Siberian populations (Derbeneva *et al.,* 2002a; 2002b; Derenko *et al.,* 2003). In northern Siberia, among Mansis, a novel branch, a sister group of hg W, was recently characterized (Derbeneva *et al.,* 2002b). In Central Asia, less than 5% of the haplotypes belong to South Asian-specific branches of hg U2 and hg M (Comas *et al.,* 2004). These hgs have probably arisen through *in situ* diversification in early Upper Palaeolithic (Kivisild *et al.,* 1999a; Bamshad *et al.,* 2001).

The derived lineage groups of hg M – hgs D (D4 and D5), G (G1 and G2), C, Z; and those of hg N – hgs A and Y form the majority of the mtDNA pool in northern and northeastern Asia (Torroni *et al.,* 1993a; Starikovskaya *et al.,* 1998; Schurr *et al.,* 1999; Kivisild *et al.,* 2002; Derenko *et al.,* 2003; Fedorova *et al.,* 2003; Puzyrev *et al.,* 2003) and are common in East Asia (Kivisild *et al.,* 2002; Yao *et al.,* 2002a; 2002b). These hgs are predominant eastern Eurasian-specific hgs also in the Central Asian mtDNA gene pool (Kivisild *et al.,* 2002; Comas *et al.,* 2004; Quintana-Murci *et al.,* 2004). The spread of hgs C, D, and G shows a decreasing gradient towards South and Southeast Asia. These hgs are rare in most Southeast Asian populations (groups L, R and K, respectively, in Ballinger *et al.,* 1992; Schurr and Wallace, 2002). However, hg D, which is common in East Asia and Siberia, is rather unevenly spread in Southeast Asia – it is rare or absent among Vietnamese, Malay and Sabah Aborigines (group L in Ballinger *et al.,* 1992), but found at moderate frequencies (17%), for example, in Thailand (Fucharoen *et al.,* 2001; Oota *et al.,* 2001; Yao et al. 2002). Some sub-branches of hgs D4 and G2 might have had their earliest diversification in Central Asia (Comas *et al.,* 2004). Hgs A and Y have phylogeographic patterns, similar to hgs C and D – these hgs are more frequent in northern Asian populations (Starikovskaya *et al* 1998; Schurr *et al.,* 1999; Saillard *et al.,* 2000) and are, likewise, virtually absent in South Asia (Kivisild *et al.,* 1999a,b; 2003b; Bamshad *et al.,* 2001) and rare in Southeast Asia (Fucharoen *et al.,* 2001; Oota *et al.,* 2001; Schurr and Wallace, 2002; Yao *et al.,* 2002a). Hg Z, the sister clade of C, stemming from the common node M8 (Kivisild *et al.,* 2002), which was first described in north Siberians (Schurr *et al.,* 1999), is, interestingly, present also in North European populations, among Finns (Meinilä *et al.,* 2001) and Saami (Sajantila and Pääbo, 1995; Dupuy and Olaisen, 1996; Lahermo *et al.,* 1996; Delghandi *et al.,* 1998), where the hg Z1 haplotypes spread in northern Eurasian populations (with HVS-I motif 16129-16185-16223-16224-16260-16298) can be found. In those populations hg D5b haplotypes with HVS-I transitions at nps 16126-16136-16189-16223-16390, otherwise found from southern Siberian populations (Derenko *et al.,* 2003), can be observed (Meinilä *et al.,* 2001).

South Asia, with its own specific branches of M and N, represents the third mtDNA domain in Asia. Hgs M2, R5, and subgroups U2a, U2b, and U2c of U2 (Kivisild *et al.,* 1999a; 20003; Quintana-Murci *et al.,* 2004), which make up more than 15% of South Asia mtDNAs, each show coalescent times over 50 KYA (Metspalu *et al.,* 2004). These hgs form a set of the most ancient Indian-specific hgs identified so far. A number of novel Indian-specific basal N and R lineages

(N5, R7, R8, R30 and R31) were recently identified from complete sequences (Palanichamy *et al.,* 2004). A large and rather superficially characterized variety of sub-branches of hg M including M2-M6, are spread at high frequencies only in South Asia (Kivisild *et al.,* 1999a; 2003b; Bamshad *et al.,* 2001). Hgs M7, M8 (other than C and Z), M9 and N9a are common in East and Southeast Asian populations (Fucharoen *et al.,* 2001; Oota *et al.,* 2001; Kivisild *et al.,* 2002; Yao *et al.,* 2002a; 2002b; Kong *et al.,* 2003a; 2003b), and some of those (M7 and M9) have also been found occasionally from some Siberian populations (Derbeneva *et al.,* 2002b; Derenko *et al.,* 2003). The Southeast Asia mtDNA pool consists mostly of hg B and hg R9, the latter includes also hg F. These haplogroups might have had their earliest diversification in Southern China and/or Southeast Asia (Yao *et al.,* 2002b). In some populations, like in Polynesians, hg B is almost the only variant of mtDNA (haplotypes with 9bp deletion) (Ballinger *et al.,* 1992; Redd *et al.,* 1995; Sykes *et al.,* 1995; Lum and Cann, 1998). In most of the Siberian populations its frequency is low (Derenko *et al.,* 1999; 2003; Fedorova *et al.,* 2003).

2.6.2. mtDNA variation in Europe

The analysis of classical genetic markers has shown that Europe as a whole is quite homogeneous – the genetic distances between different populations are relatively short and the genetic landscape is rather uniform (Di Rienzo and Wilson, 1991; Piazza, 1993; Cavalli-Sforza *et al.,* 1994). Only some clear outliers like the Saami, Sardinians, Basques and some others have been shown to emerge from this homogeneous entity (Cavalli-Sforza *et al.,* 1994).
More than 90% of maternal lineages present in European populations can be classified into eight major haplogroups (H, V, T, J, N1, U, X, W; see figure 4), characteristic to western Eurasians in general (*e.g.* Torroni *et al.,* 1994d; 1996; Richards *et al.,* 1998; 2000; Macaulay *et al.,* 1999b; see also table 1). All of them coalesce to the common node N in mtDNA phylogeny (figure 3). As it was already indicated above, only a minor part of maternal lineages that have been found in Europe belong to East or South Asian-specific haplogroups. Yet it is worthwhile to note here that the uniformity of the genetic landscape of the distribution of mtDNA haplogroups in Europe is a term should not be used in its absolute meaning. It has already been shown (Richards *et al.,* 2002; Richards, 2003) that a deeper phylogenetic analysis allows the revealing of significant differences in the spread of mtDNA sub-haplogroups in Europe – and not only among genetic outliers.
The phylogenetic classification of European mtDNA relies mostly on the combined usage of diagnostic coding region polymorphisms and sequence information of both hypervariable segments of mtDNA. Further refinement of the present nomenclature, due to higher phylogenetic resolution, can be achieved by the complete sequencing of mtDNA genome (*e.g.* Finnilä *et al.,* 2001; Herrnstadt *et al.,* 2002). The majority of European-specific haplogroups (HV, TJ and U) stem from a large nested lineage cluster R of the macrohaplogroup N (see figure 4).
Hg H is by far the most frequent maternal lineage cluster in Europe. Its frequency is the highest (40-60%) in western and northern European populations, but it is also common (20-30%) in the populations of the Near East (Torroni *et al.,* 1998; Richards *et al.,* 2000) and is well visible in North Africans (Corte-Real *et al.,* 1996; Rando *et al.,* 1998; Stevanovitch *et al.,* 2004) and Central Asians (Metspalu *et al.,* 1999; Comas *et al.,* 2004; Quintana-Murci *et al.,* 2004), less so in South Asians (Passarino *et al.,* 1996; Kivisild *et al.,* 1999a; 1999b; 2003b, Bamshad *et al.,* 2001) and in Native Siberians (Torroni *et al.,* 1998; Derbeneva *et al.,* 2002a; 2002b; Derenko *et al.,* 2003; Fedorova *et al.,* 2003). It is believed that the initial expansion of hg H took place most probably in the Near East about 25,000 YBP (Richards *et al.,* 2000). Hg H includes the Cambridge Reference Sequence (CRS, Anderson *et al.,* 1981), which is one of the most frequent HVS-I haplotypes in Europe and occupies the central node in the hg H topology, provided additional coding region mutations are not considered.
There have been several attempts to refine the inner structure of H (*e.g.* Helgason *et al.,* 2001; Malyarchuk and Derenko, 2001) to be able to trace the phylogeographical spread of different

branches of this major haplogroup. Based on mtDNA complete sequences, seven sub-branches of hg H – H1 and H2 (Finnilä *et al.,* 2001) along with H3 and H4 (Herrnstadt *et al.,* 2002) and H5-H7 (Quintans *et al.,* 2004) have been defined.

A broad phylogeographic analysis and refinement of hg H sub-clades has been performed by Loogväli *et al.,* (2004), thus filling a gap in previous insufficient depth and width of the phylogenetic analysis of the predominant hg among European populations. Making use of the coding sequence information from 267 mtDNA hg H sequences, Loogväli and colleagues (2004) have analyzed 830 mtDNA genomes, from 11 European, Near and Middle Eastern, Central Asian and Altaian populations. In addition to the seven previously specified sub-haplogroups, the presence of the 15 novel sub-clades of hg H were determined in the extant human populations of western Eurasia. The refinement of the phylogenetic resolution has allowed to resolve a large number of homoplasies in phylogenetic trees of hg H based on the HVS-I of mtDNA. As many as 50 out of 125 polymorphic positions in HVS-I were found to be mutated in more than one sub-cluster of hg H. The phylogeographic analysis revealed that sub-haplogroups (sub-hgs) H1*, H1b, H1f, H2a, H3, H6a, H6b and H8 demonstrate distinct phylogeographic patterns.

The largest sub-cluster is sub-hg H1, which comprises about 30% of hg H, and 13% of the total European mtDNA pool. H1 is most frequent in the Iberian Peninsula covering about 46% of local hg H lineages (Pereira *et al.,* 2004; Quintans *et al.,* 2004). In the Near East the frequency of H1 does not exceed 6% and its relative frequency in respect to hg H is lower than that seen in Europe, 14%. In the Central Asian populations, where hg H makes up about 11% of the local mtDNA pool, only 6% of H samples belong to sub-hg H1. Sub-hg H1b is found throughout the area of the spread of hg H, being more frequent in Eastern Europe and North Central Europe (about 7% and 5% of hg H, respectively). It was also found to make up about 5% of hg H in Siberian Mansis. A minor sub-hg H1f constitutes a quarter of the selected subset of Finnish hg H genomes of Finnilä, Lehtonen and Majamaa (2001), being almost absent elsewhere in Europe. Confirmation of the high frequency of this rare variant of mtDNA among northern-central Finns, characterized by HVS-I motif 16093-16189, can be found in the Finnish data of Meinilä, Finnilä and Majamaa (2001), reflecting founder effects in the Finnish population history (de la Chapelle and Wright, 1998; Kittles *et al.,* 1999; Peltonen, Palotie and Lange, 2000). Like H1b, sub-hg H2a occurs more frequently in eastern than in western European hg H genomes, 6.5% and 1.1%, respectively, when averaged over populations. In contrast, sub-hg H3 was found to be more frequent in western (11.7%) than in eastern European hg H pool (4.1%) and is virtually absent in Anatolia and in the Near East. The high frequency of mtDNA hg H3 extends to the Iberian Peninsula, where H3 constitutes about 17% of hg H and is the highest detected so far (Pereira *et al.,* 2004; Quintans *et al.,* 2004). The coalescence ages of H2a1 and H3 fall to the period of postglacial recolonization in Europe (11,220±5,000 YBP and 11,100±1,600, respectively), suggested first for mtDNA hg V (Torroni *et al.,* 1998; 2001a). The Near Eastern samples cluster together with Central Asian mtDNAs in the sub-hgs H6b and H8, which are very rare in Europe. The finding is demonstrating a separate flow of maternal lineages south of the Caspian and the Black Sea in addition to well-known longlasting migrations of pastoral nomads alongside the steppe belt that connects the Danube Basin, over the Pontic-Caspian, with Central Asia, Altay and Manchuria. In contrast to that found in Europeans, sub-hgs H6 and H8 among Central Asian/Altaian populations are characterized by distinctly divergent haplotypes. This finding may reflect a long-time separation of Asian and European H6 and H8 mtDNA pools and/or an earlier expansion of H6 in the eastern part of its present range. Indeed, the coalescence age of H6 in Central Asians is very deep – 40,400±16,400 YBP (Loogväli *et al.,* 2004).

The commonly used HVS-I clock (Forster *et al.,* 1996) places the initial expansion of hg H in the Near East to about 23,000 to 28,000 YBP (Richards *et al.,* 2000). The ancestral clades of hg H, pre-HV and HV* have their combined present range predominantly in the Near and Middle East, and in the Caucasus (Metspalu *et al.,* 1999; Richards *et al.,* 2002), implying this could have been the region where the pre-HV/HV clade started to diversify and, possibly, where the earliest hg H

variants may have first appeared. However, most sub-clusters of hg H exhibit coalescence ages, corresponding to the beginning of their expansion in the Late Upper Paleolithic. In this respect obtained results support an earlier proposition that hg H was the major mtDNA haplogroup participating in the recolonization of Europe after the Last Glacial Maximum (Torroni *et al.,* 1998; Richards *et al.,* 2000). Sub-hgs H1 and H3 have their highest frequencies in the Iberian Peninsula. These sub-hgs may have been the companions of mtDNA Hg V in the post-glacial re-peopling of Europe from a refuge area in Iberia (Torroni *et al.,* 1998). However, in contrast to hg V, suggested coalescence ages of H1 and H3, 13,400±3,000 and 8,600±2800 YBP, respectively (Pereira *et al.,* 2004), do not imply deeper phylogeny of H1 and H3 in Iberia compared to the rest of Europe (review from Loogväli *et al.,* 2004). Later, Achhili *et al.,* (2004) have also showed that sub-hgs H1 and H3 display frequency peaks, centred in Iberia and surrounding areas, with distributions declining toward the northeast and southeast—a pattern extremely similar to that previously reported for mtDNA haplogroup V (Torroni *et al.,* 1998). However, the calculated coalescence ages of H1 and H3 (~11,000 years) are close to that previously reported for V. These findings have major implications for the origin of Europeans, since they attest that the Franco-Cantabrian refuge area was indeed the source of late-glacial expansions of hunter-gatherers that repopulated much of Central and Northern Europe from ~15,000 years ago.

Described above results demonstrate that a seemingly uniform spread of this major human mtDNA clade in western Eurasian populations hides within itself a complex structure of phylogeographically informative sub-clades. However, it is evident that additional knowledge at the level of complete mtDNA sequences is still needed for a truly comprehensive cataloguing of hg H diversity, in particular more effectively covering its variation in the Mediterranean, Near and Middle Eastern and Central Asian/Altaian populations. Nevertheless, most of the present-day Near Eastern and Caucasus area as well as European variants of hg H started to expand after the Last Glacial Maximum (LGM) and presumably before the Holocene. Yet importantly, several hg H sub-clades in Near East and Southern Caucasus region coalesce to the pre-LGM period. Furthermore, irrespective of their common origin, significant differences between the distribution of hg H sub-hgs in Europe and in the Near East and South Caucasus imply limited post-LGM maternal gene flow between these regions. In a contrast, the North Caucasus mitochondrial gene pool has received an influx of hg H variants, arriving from the Pontocaspian/East European area (Roostalu *et al.,* 2007).

Phylogenetically closely related clades to hg H – (preHV)1, HV1 and HV2 – are present predominantly in the Near East, Middle East and in the Caucasus (Metspalu *et al.,* 1999; Richards *et al.,* 2002; 2003). Hg V is the sister-clade of hg H (Torroni *et al.,* 1998). The frequency of hg V does not exceed the level of 1-6% in the mtDNA gene pool of most European populations. As exceptions, the Basques and Catalans in the Iberian Peninsula can be named, among whom the frequency of hg V is more than 20%. A particular exception is also the Saami in the Scandinavian Peninsula. Among the latter, the frequency of HVS-I haplotypes carrying characteristic to hg V transition in np 16298 has been shown to be especially high, ranging over 40% (Sajantila *et al.,* 1995). The diversity of hg V is, however, very low among the Saami (Torroni *et al.,* 1998, Tambets *et al.,* 2004). It has been proposed that hg V is a mtDNA marker for the population expansion after LGM that started from the Iberian Peninsula 10,000-15,000 years ago (Torroni *et al.,* 1998).

Hg U is the most ancient haplogroup in Europe and embraces numerous phylogeographically different sub-clades (figure 4), some of which can be found from Africa (*e.g.* Rando *et al.,* 1998; Rosa *et al.,* 2004), Siberia (*e.g.* Derbeneva *et al.,* 2002b; Derenko *et al.,* 2003; Fedorova *et al.,* 2003), the Near East and the Caucasus-area (Metspalu *et al.,* 1999; 2004; Richards *et al.,* 2003), as well as in South Asia (Kivisild *et al.,* 1999a; 1999b; 2003b; Bamshad *et al.,* 2001). The expansion of hg U has likely started more than 50,000 YBP (Torroni *et al.,* 1996; Richards *et al.,* 1998; 2000). Hg U5 is the largest and most diverse branch of hg U, and its coalescent age is 45,000-53,000 YBP. It can be further divided into a number of subclades (Richards *et al.,* 1998; 2000), but because many of them are thus far defined only by fast evolving nps in HVS-I, more

extensive knowledge about its coding region variation is needed before in-depth phylogenetic analysis of different subclades can be vigorously applied. Tambets *et al.,* (2003) have performed the analysis of a large dataset of U5 mtDNAs (n=526) based mostly on HVS-I sequences and essential diagnostic site for hg U (+12,308 *Hinf*I) as well as on coding region information from Finnilä *et al.,* 2001 and Herrnstadt *et al.,* 2002. The study revealed that U5 ramifies into many potential sub-founders (U5a and its sub-hgs: U5a1, U5a1a, U5a1b; U5b and its sub-hgs: U5b1, U5b1b, U5b1a, U5b2). The important aspect of this analysis was that almost all of them show a star-like topology and exhibit post-LGM coalescence ages around the end of Pleistocene period, suggesting that hg U5 underwent a post-LGM expansion phase. However, some of the sub-clades of hg U5 can be recognized also by their specific HVS-I motifs. One of these is the so-called "Saami-motif", distinguished from the root of hg U and the CRS by three transitions in nps 16189, 16144 and 16270 (Sajantila *et al.,* 1995). In contrast, Tambets *et al.,* (2004) analyzed the informative coding region nps in U5b topology and showed that "the Saami variant" of U5b (U5b1b1), predominant among the Saami, is widely spread among different eastern European populations, extending , at low frequencies, also to western Europe and to the Caucasus.

Hg U4 is the second largest sub-clade of U in Europe. According to Richards *et al.,* (1998), hg U4 dates back to more than 25,000 years and, similarly to hg U5, started to expand before the LGM. Its frequency seems to be higher in eastern and southern than in western European populations and is the highest (16%), surprisingly, among western Siberian Ugric-speaking Mansis (Derbeneva *et al.,* 2002b). The phylogenetic tree of U4 that was constructed from hg U4 mtDNAs (n=386) showed a limited number of subfounders. Sub-clade U4a (defined my HVS-I motif 16134-16356) and U4b (defined my HVS-I motif 16179-16356) likely encompass monophyletic lineages. Lineage cluster U4c (defined my HVS-I motif 16356-16362) is more likely a paraphyletic group because np 16362 is one of mutational "hotspots". Based on the coalescence ages for different subfounders in different geographical subsets of U4, have been found that in eastern Europe, the coalescence ages are more or less the same for all main founder-haplotypes (except for U4b, that is very rare in eastern regions of Europe), being around the LGM (20,000-22,000 YBP) (Tambets *et al.,* 2003). Taking into account the data from other disciplines (*e.g.* Dolukhanov, 2000), this might suggest that the expansion of U4 in eastern Europe started during formation of the Periglacial (northern Ukrainian) refugium around the peak of the LGM.

Hgs U1, U2, U3, U6, U7 and U8 are, particularly in Europe, less frequent subclades of hg U. From those, U1, U3, U6 and U7 are present at relatively higher frequencies in southern European populations and in the Near East, North Africa (Macaulay *et al.,* 1999b; Metspalu *et al.,* 1999; 2004; Tambets *et al.,* 2000), whereas U2 can be found throughout the Near East and Europe (Macaulay *et al.,* 1999b). As it was indicated above, distinct sub-branches of U2 are specific for South Asia (Kivisild *et al.,* 1999a; 2003b; Bamshad *et al.,* 2001; Metspalu *et al.,* 2004). Hg K, one of the first described haplogroups in the European population (Torroni *et al*. 1994d), was shown to belong to hg U (Hofmann *et al.,* 1997). Hg K seems to be more common in western than in eastern European populations (table 1), and is well present in the Near East and Anatolia (Kivisild *et al.,* 2003a). The highest frequency of hg K has been observed among Ashkenazi Jews (30%, Ritte *et al.,* 1993; Behar *et al.,* 2004).

Hgs T and J (figure 4) are sister clades, present in European populations at similar frequencies (6%-14%, see also table 1). Both of them have a complex inner topology with several sub-founders (*e.g.* Finnilä and Majamaa, 2001; Herrnstadt *et al.,* 2002), some of which might have arrived to Europe with the Neolithic immigration from the Near East (Richards *et al.,* 2000).

Hgs that stem directly from node N (I, N1a, N1b, N1c, W, X, see figure 4) are relatively rare in Europe and do not usually exceed the level of 5% there (Richards *et al.,* 1998). Hg I, together with N1a, N1b and N1c, belongs to the clade N1 (see figure 4) and is spread mostly in northern and western Europe (Richards et al. 2000). Hg X, despite its minor contribution to the package of maternal lineages of most of the populations, has a surprisingly wide geographical distribution. Described first as a European specific mtDNA variant (Torroni *et al.,* 1996) it has by now been

found from North and East Africa, Near East, Siberia as well as among Native Americans (Forster *et al.,* 1996; Brown *et al.,* 1998; Smith *et al.,* 1999; Derenko *et al.,* 2001; Reidla *et al.,* 2003). Hg W has higher diversity in southern than in northern European populations (Richards *et al.,* 1998) and it has been found also, for example, in South Asia (Kivisild *et al.,* 1999b; Metspalu *et al.,* 2004). The frequency of hg W is particularly high among Finns (table 1), its diversity, however, is very low there (Meinilä *et al.,* 2001).
Most European haplogroups show greater haplotype diversities as well as deeper coalescence ages in the Near East than in Europe (Torroni *et al.,* 1998; Richards *et al.,* 2000), supporting their Near Eastern origin. The coalescence ages of the primary founder haplotypes have shown that the majority of them have migrated to Europe in the Late Upper Palaeolithic-Mesolithic period (Figure 5) (Richards *et al.,* 2000). One exception is hg V, which seems to have originated and expanded within Europe 10,000-15,000 YBP (Torroni *et al.,* 1998). Another is hg U5. Although it occurs at about 2% in the Near East, its phylogeographical spread suggests that it evolved mainly within Europe during the past 50,000 years and its presence in the Near East can be result of back-migration from Europe (Richards *et al.,* 2000).

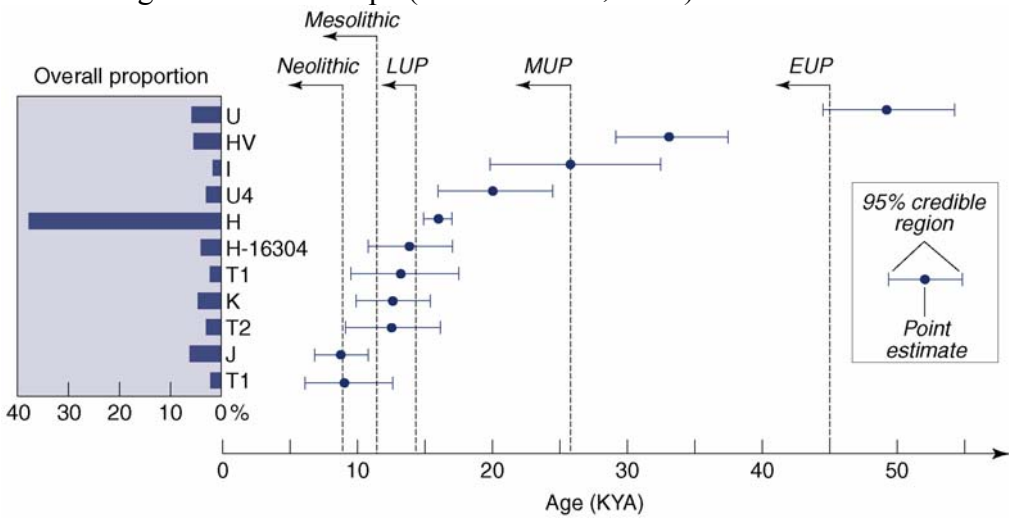


Figure 5. Estimated ages of major European mtDNA haplogroup founders (Jobling, Hurles and Tyler-Smith, 2004). Ages are taken from Richards et al. (2000), and were calculated using the statistic ρ. Proportions of each haplogroup in Europe as a whole are given in the bar chart to the left. EUP, MUP, LUP correspond to Early, Middle and Late Upper Palaeolithic; KYA corresponds to thousands years ago or kilo years ago.

2.7. Pre-historical, archaeological and linguistic context for present European mtDNA variation

The beginning of the Upper Palaeolithic, 40,000-50,000 YBP, marks the first appearance of AMHs in Europe. During the peak of the Last Ice Age, about 24,000-20,000 YBP, a large part of Europe was unsuitable for human occupation. In the cold period the population of Europe was concentrated to the so-called refuge areas or refugia – regions that maintained more favourable environmental conditions at the LGM-period. After the LGM, northern regions of Europe gradually re-colonized by Late Upper Palaeolithic populations spreading out from different refugia – the Franco-Cantabrian, Ukrainian and possibly some others, situated, *e.g.* in the southern Alps, in so far less well-identified locations south of the lower Danube and Balkans (Housley *et al.,* 1997; Dolukhanov, 2000). One still hotly debated issue on the possible origin of modern Europeans is associated with the onset of farming, its spread in Europe and its genetic influence to the European Mesolithic populations. To competing hypotheses – the demic and cultural diffusion models – have been presented. According to the demic diffusion model, food production let to a slow expansion of Neolithic source population from the Near East due to the

population growth, resulting from the advantages of agriculture ("wave of advance", Ammerman and Cavalli-Sforza, 1984; Cavall-Sforza *et al.,* 1994).
According to the most extreme views, it led to the total replacement of less numerous Mesolithic hunter-gathers, who had lived in Europe before (Barbujani *et al.,* 1990; Chikhi *et al.,* 1998; 2002). Yet the classical presentation assumes that the demic diffusion involved a substantial minority of Neolithic newcomers (*ca* 27%), arriving from Near East – either from Levant or Anatolia (Cavalli-Sforza *et al.,* 1994). The model of demic diffusion was embraced also by some linguists (Renfrew, 1987) and it was proposed that: 1) the genes of Anatolian and Near Eastern populations; 2) the technology of farming; and 3) Indo-European languages, spoken today in most regions of Europe, were together brought to Europe in the course of the same migration. However, most linguists did not agree with this model and preferred the theory of Marija Gimbutas (1970), according to which Proto-Indo-Europeans had spread west with Bronze-Age "Kurgan culture" of the eastern European steppe (Mallory, 1989; Hines, 1991). More recently, and extensive study on 87 languages yielded a support for the Anatolian theory of Indo-European origin (Gray and Atkinson, 2003). The study also showed evidence of a period of rapid divergence giving rise to the Italic, Celtic, Balto-Slavic and Indo-Iranian families in a close time frame to that suggested for a possible Kurgan expansion. Thus, these two linguistic hypotheses need not be mutually exclusive. Several theories have been postulated also on the formation of the Finno-Ugric branch of the Uralic language family, spoken by a minority of inhabitants of Europe, like, for example, by the Finns, Karelians, Estonians and the Saami. The earliest hypotheses associated the arrival of Finno-Ugric speakers from the Volga-Oka River region, which was assumed to be the Uralic homeland, in Europe *ca* 6,000 YBP with Neolithic Combed Ware culture. According to the latest hypotheses, the migration of Finno-Ugric speakers occurred from the south, almost simultaneously with the retreat of the ice sheets (*e.g.* Wiik, 2000; Poikalainen, 2001). The opposite model of the origin of modern Europeans assumes, on the contrary, that the onset of farming took place not because of any significant migration of people from the Near East, but because of a cultural transition in the form of diffusion of ideas (Barker, 1985; Whittle, 1996). One intermediate model, the pioneer colonization, assumes the selective migration of only small groups from western Asia (Zvelebil, 1986, 2000, 2001; van Andel, 2000) and the continuity of indigenous Mesolithic populations. This last model, with some modifications (Wilson *et al.,* 2001), explaining the presence of the Neolithic component in comparable frequencies to this of Central and South Europe, also in North and northeastern Europe, has also gained support from mtDNA studies (Richards *et al.,* 1996; 2000; Torroni *et al.,* 1998; Richards, 2003), where it was suggested that the most of the contemporary maternal lineages in Europe have their ancestry in the Late Glacial expansions within Europe, associated with the climatic improvements following the Last Ice Age, whereas only a minor part is dated either to the initial early Upper Palaeolithic settlement of the continent by AMHs (hg U), or brought to Europe by later immigrants during the early Neolithic (hgs J, T1, U3, few subclusters of H and W, Richards *et al.,* 2000).

2.7.1. Archaeological, linguistic, biochemical and genetic evidence of the peopling of Latvia

The first clear evidence about the existence of genetic differences between individuals was obtained more than a century ago when Karl Landsteiner in 1901 (Landsteiner, 1901) described different blood groups of the ABO-system. More extensive studies in this area started in the 1950s-1960s when systematic analysis of the variation of proteins in different populations became possible. The transition from the analysis of protein polymorphisms to the studies of diversity of genes started in 1980s. While the main aim of the first studies was to get information about how genetic variation is associated with diseases, they also widened the understanding of how this variation may reflect demographic history of humans.
Genetic investigations of classical markers (*e.g.* TF DCHI, PI, LW^b) among Latvians starting to emerge in 1990s have revealed genetic stratification at the intra-population level (Beckman *et*

al., 1999; Krumina *et al.,* 2001), as well as differences in the Latvians compared to other Indo-European and Finno-Ugric speaking populations of the Baltic Sea region (Beckman *et al.,* 1998; Sistonen *et al.,* 1999). Several examples of the population genetics studies performed on the Latvian population are provided herein.

The ABO blood group allele B, with a high frequency in Asia, is a well established marker of eastern influence in Europe. The frequency of the B allele is about twice as high in all populations on the eastern side of the Baltic Sea compared to that in Scandinavia indicating a definite eastern influence (Mourant *et al.,* 1976). Increased frequencies of the B allele have also been found in eastern mainland Sweden and on the island of Gotland suggesting an eastern (Finno-Ugric or Baltic) influence in these areas (Beckman *et al.,* 1959). In an attempt to trace Finno-Ugric genetic influence and migrations, Beckman *et al.,* (1998) studied the distribution of transferrin alleles in populations of the Baltic Sea region. Significant regional differences were found in transferrin variants typical of Finno-Ugric populations, particularly TF DCHI. This variant was found in Finland and Estonia and on the island of Gotland, but not among Baltic peoples (Latvians and Lithuanians) and in southern mainland Sweden. The finding of TF DCHI and also another Finno-Ugric transferrin allele (TF DFIN) in the Gotland population indicates an eastern (Estonian, Livonian or Finnish) influence on this island. Data from previous investigations also have indicated that the populations east of the Baltic Sea showed a mixture of eastern and western genetic characteristics (Mourant *et al.,* 1976; Viikma and Heapost, 1996) , *e.g.* the high B-allele frequency, an eastern trait, is combined with a high frequency of the Rh-negative blood group, a typically West-European trait. Sistonen *et al.,* (1999) have studied the Baltic populations with respect to exceptional variations in the frequencies of the Landsteiner-Wiener (LW) blood group. The frequency of the uncommon LW^b gene was high in the Balts, around 6% among Latvians and Lithuanians, very low among the other western Europeans (0–0.1%) and apparently absent in Asiatic and African populations. From the Baltic region of peak frequency there was a regular decline of LW^b incidence (a descending cline) in the neighbouring populations: 4.0% in the Estonians, 2.9% in the Finns, 2.2% in the Vologda Russians, and 2.0% in the Poles. Thus, the distribution of LW^b suggests considerable and extensive Baltic admixture, especially in the north and northeast direction. In Southern Sweden with an LW^b frequency of 0.3%, the Baltic influence appeared slight, while in the population of the Swedish island Gotland in the middle of the Baltic Sea there was a significantly increased LW^b frequency of 1.0% compared with that of Western European countries. Based on population data, it is plausible that the expansion of this point mutation occurred only once during human history. Furthermore, Sistonen *et al.,* (1999) have concluded that the expansion of the LW^b mutation occurred in Balts and that LW^b can be considered a "Baltic tribal marker", its presence in other populations being an indicator of the degree of Baltic genetic influence.

The analysis of one of the most common genetic diseases in Europeans – phenylketonuria (PKU) – has revealed the predominance of a single mutation in the phenylalanine hydroxylase (*PAH*) gene, R408W of haplotype 2, in the Baltic states. Although well spread throughout Eastern Europe, R408W has its frequency peak in Latvia, Lithuania and Estonia, there comprising about four fifth of PKU haplotypes (Lilleväli *et al.,* 1996; Kasnauskiene *et al.,* 2003; Pronina *et al.,* 2003). It has been suggested that this mutation originated in an ancient eastern European population, from where it spread westward (Eisensmith *et al.,* 1995).

Alpha 1- antitrypsin (AAT) deficiency is a genetic disorder characterized by a low serum level of ATT and a high risk of pulmonary emphysema and a liver disease at a young age (Cox, 1995). ATT deficiency is one of the most common inherited diseases in Caucasians. Allelic frequency for the most common protease inhibitor (PI) Z mutation is 1-2% in Caucasians of Northern Europe descent (Lieberman and Sastre, 1986; Beckman *et al.,* 1999). The distribution of alpha 1-antitrypsin (PI) alleles was studied in an attempt to elucidate migrations and admixture between populations in the Baltic Sea region. The frequency of the PI Z allele, a typically North-Western European marker gene, showed a highly significant regional variation in the Baltic Sea region.

The highest frequency (4.5%) was found in the western part of Latvia (Courland). The PI S allele, another marker of West-European influence, also showed an increased frequency in the Courland population. These results indicate that among the populations east of the Baltic Sea the Curonian population has the most pronounced West-European influence. Increased frequencies of the PI Z alleles and S alleles in Courland may have been caused by migrations from mainland Sweden and the island of Gotland (Beckman *et al.,* 1999). Based on the analysis of 251 healthy individuals who had lived for at least three generations in selected regions of Latvia, the frequency of PI Z allele in ethnic Latvians was calculated to be 1:303. Thus this represents the highest population frequency of Z allele in the Caucasian populations reported so far (Maliseva-Lace *et al.,* 2004). Recently, Blanco *et al.,* (2006) have performed a large scale population studies on estimates of the numbers of individuals carrying the two most common deficiency alleles, PI S and PI Z, for alpha1-antitrypsin deficiency in Europe. A total of 75,390 individuals were selected from 21 European countries (including the Latvian population). The largest number of ZZ (5,000-15,000) were in Italy, Spain, Germany, France, the UK, Latvia, Sweden and Denmark, followed by Belgium, Portugal, Serbia-Montenegro, Russia, The Netherlands, Norway and Austria (1,000-2,000), with < 1,000 in each of the remaining countries. A remarkable lack in number of reliable epidemiological studies and marked differences among these European countries and regions within a given country was also found. The human leukocyte antigen system (HLA) has been extensively studied from population genetics and evolutionary perspectives. Moreover, population and disease studies emphasize consideration of ethnic variation in human populations. The HLA region, the major histocompatibility complex (MHC) of humans, contains a number of closely linked, highly polymorphic genes whose products control a variety of functions concerned with the regulation of immune responses. Over 50 diseases have been shown to be associated with the HLA region, these include: autoimmune diseases such as insulin dependent diabetes and multiple sclerosis; cancers such as Hodgkin disease; and infectious diseases, such as tuberculosis and AIDS. These diseases are genetically complex and involve genetic heterogeneity in the HLA region and non-HLA genes (Thomson, 1991). Population data studies for HLA class-I (B allele frequencies) and class-II two-locus haplotypes (DR and DQ loci) were carried out in the Latvian population as well as examinations for alleles and haplotypes distribution in patients with rheumatic heart disease (RHD), rheumatic fever (RF) and Insulin-Dependent Diabetes Mellitus (IDDM, or Diabetes mellitus type 1). HLA genotyping showed that HLA class II DRB1*07-DQB1*0401-2 and DRB1*07-DQB1*0302 could be the risk alleles and HLA class II DRB1*06 and DQB1*0602-8, the protective ones in the case of RHD and RF syndromes (Stanevicha *et al.,* 2003). The genes encoding the HLA-DQ heterodimer molecules, DQB1 and DQA1, have been found to have the strongest association with IDDM risk, although there is cumulative evidence for the effect of other gene loci within the major histocompatibility complex gene region. Nejentsev *et al.,* (1997) have analyzed at the population level the effect of DR4 subtypes and class I, HLA-B alleles, on IDDM risk when the influence of the DQ locus was stratified. In all three populations studied (Estonian, Latvian, and Russian), DQB1*0302 haplotypes most frequently carried DRB1*0401 or DRB1*0404. DRB1*0401 was the most prevalent subtype in IDDM patients, whereas DRB1*0404 was decreased in frequency. When HLA-B alleles were analyzed, strong associations between the presence of specific B alleles and DRB1*04 subtypes were detected. The HLA-B39 allele was found significantly more often in DRB1*0404-DQB1*0302-positive patients than in healthy control subjects positive for this haplotype. The observed results have demonstrated that DQ and DR genes cannot explain all of the HLA-linked susceptibility to IDDM, and that the existence of a susceptibility locus telomeric to DR is probable. Another survey showed that the rare HLA-DQB1*0304 allele was found increased among IDDM patients in the populations of the eastern Baltic region. Its frequency among IDDM patients was 4.5% compared to 1.1% in healthy controls in the combined series of Estonian, Latvian and St. Petersburg Russian populations (P=0.0001). HLA-DQB1*0304 in these populations was associated with DRB1*0408, and the haplotype was further characterized by a B35 allele and a typical combination of microsatellite

markers from the TNF gene region. The result is compatible with the significance of the 57th amino acid in the DQ beta-chain but also emphasizes the importance of alleles in other HLA loci adjacent to DQ in the determination of IDDM susceptibility (Ilonen *et al.,* 1997).

Initial anthropological, archaeological, and linguistic studies of Latvians were performed in the end of the 19th and in the early period of the 20th centuries and were addressed to the origin and spread of Baltic-speaking populations as well as about the probable genetic influence of other tribes, like Finno-Ugric- and Slavonic-speaking neighbours, surrounding the region during prehistory (Waeber, 1897; Primanis, 1937). The territory of the eastern coast of the Baltic Sea was permanently settled relatively late, as the land became habitable only approximately 12 000–10 000 years ago, after the end of the last glaciation. The first inhabitants of the region were hunter-gatherers, probably carriers of the Svidrian and Magdalenian cultures (Gimbutas, 1963). These pioneer hunter-gatherers likely arrived predominantly from the major European glacial refugia, situated in the present territory of Ukraine and Francocantabria (Dolukhanov, 2000). The formation of Baltic tribes in the territory was a complex process, associated with the interaction of different population groups, with the dispersal of Indo-European languages as well as with the process of neolithization in the region.

The **Balts** or **Baltic peoples**, defined as speakers of one of the Baltic languages, a branch of the Indo-European language family, are descended from a group of Indo-European tribes who settled the area between lower Vistula and upper Daugava and Dnieper rivers on the southeast shore of the Baltic Sea. Because the thousands of lakes and swamps in this area contributed to the Balts' geographical isolation, the Baltic languages retain a number of conservative or archaic features.

The prehistoric cradle of the Baltic peoples according to archaeogenetic research and archaeological studies was the area near the Baltic sea and central Europe at the end of the Ice Age and beginning of the Mesolithic period. They spread in the area from the Baltic sea in the west to the Volga in the east. The Slavic cradle was in the Danubian - Krakowian region close to the Baltic. The Slavs entered the Dniepr region in the 6th century CE after the Avar invasion of Europe, conquering and assimilating most of the Eastern Balts. In the first centuries of the 1st millennium AD, the Baltic tribes settled the area between the Vistula and the Daugava. Their culture is easily recognizable and most probably they were the ancestors of the tribes of Western Balts (Old Prussians, Yotvingians and Galindians), as well as Eastern Balts (Lithuanians, Semigallians, Curonians and Latgalians) (figure 6) (Gimbutas, 1963; Sedovs, 2004).

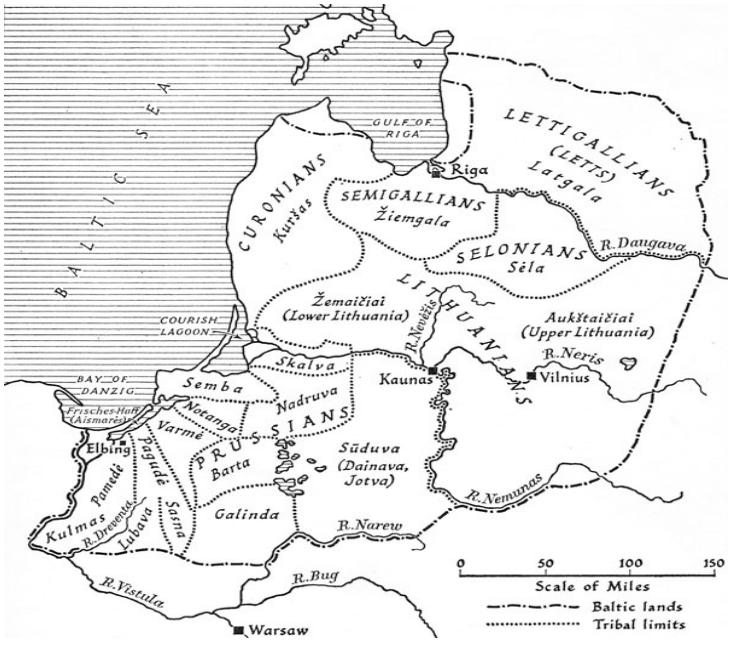


Figure 6. Baltic tribes and provinces c. A.D. 1200 (Gimbutas, 1963)

- **The most ancient population of Latvia**

The first arrival of people in the Eastern Baltic region (and in all of Northern Europe) was dependent upon the Scandinavian glacier. Its very slow retreat occurred as the result of climactic shifts which took place over the course of several millennia and were still ongoing in the late Palaeolithic period. At that time, the Scandinavian glacier still covered a fairly vast territory: the lowlands of Central Europe, Scandinavia, Finland, the Eastern Baltic region, and Northwestern Russia, including the Karelian peninsula and territories to its East. The melting of the Scandinavian glacier is divided into several climactic periods by specialists. The cold Driassian period was replaced two times by warmer periods (Belling and Allerod) which facilitated the gradual retreat and melting of the glacier.

In the population of Latvia and the rest of the Baltic region, the decisive factor was the ancient Europeans who first came into the Central European lowlands from which the glacier had retreated. This territory stretches from the Netherlands to the lower reaches of the Vistula and Niemen rivers and borders the North Sea and the Baltic Sea. Further East the territory meets the Pripet basin, as well as the upper Dnieper and Daugava rivers. The first humans arrived in the Middle European lowlands in the late Palaeolithic period, and over the course of many generations they gradually populated the farthest reaches of Northern Europe, including the Eastern Baltic. Several thousand years passed between the arrival of the first humans in the Middle European lowlands and the arrival of the first residents of the Eastern Baltic. This is particularly true of Latvia and Estonia, although the first humans arrived in southern and especially southeastern Lithuania much earlier than was the case in the rest of the Baltic region. In the settlements left behind by ancient residents in Lithuania, for example, specialist R. Rimantiene (1971) has found traces of the cultural traditions of the Lingbi, Bromme, Arensburg and Svidrian people. These ancient settlements are found mostly in southeastern Lithuania and are dated to the late Allerod and the late Driassian periods. This territory is adjacent to the Pripet basin, where archaeologists have found a large number of settlements with evidence of tool manufacturing technologies typical of the aforementioned periods. Such settlements have also been found near the Upper Dnieper, as well as at the Soża and Desna river valley (these are left bank tributaries to the Dnieper) (Kopytin 1979; Zaliznyak, 1979).

In the very last part of the Palaeolithic period, archaeologists have found, there was a migration of people of the Svidrian culture who reached Lithuania and Northwestern Russia. Archaeologists have also noted a migration of Maglemosian people during the early Mesolithic period in an eastward direction to the northwestern part of Russia. Testimony of this is provided both by settlements of these tribes which have been found in Lithuania (Rimantiene, 1971) and by the effect which the Maglemosian culture had on the territory which lies adjacent to the southwestern shores of the Oneg lake. In the early Mesolithic period, things were manufactured of bone in a method that was completely identical to that used by Maglemosian people in Denmark at the same time (Oshibkina, 1983).Thus archaeological data indicate that the migration of the Maglemosian people concluded several chronologically successive migrations of late Palaeolithic peoples from the West to the East. This suggests that at the end of that period, a genetic fund was being established in populations resident in the territory that is southwest of Latvia – the Pripet basin and the Upper Dnieper valley. This genetic fund was part of larger genetic system of late Paleolithic residents in Northern Europe. For that reason, people who settled on lands around the Upper Dnieper and the Upper Daugava during the late Palaeolithic period had close genetic links to the most ancient populations of the Middle European lowlands. That could mean that during the Mesolithic period, an anthropologically similar group of peoples lived from the Netherlands in the West to the Middle Russian highlands to the East. Local residents may have been possessed of the morphological elements of ancient Northern European peoples, whose roots were linked to the late Palaeolithic populations of Europe.

Even though the Niemen and the Daugava are separated only by a few hundred kilometers of dry land, the territory between the two rivers remained uninhabited for quite a long time. The most

ancient settlements in southern Lithuania are some 2,000 years older than the first settlements on the shores of the Daugava. Moreover, the first residents in Latvia arrived not through Lithuania, as would seem logical, but rather from the Southeast, using the Dnieper river and the Upper Daugava for this purpose. Evidence of this is provided by the fact that the most ancient settlements in Latvia are found mostly in the eastern part of the country – near the Dviete river, along the Aiviekste river, near Lake Lubāns, and on the shores of the Daugava (at Sēlpils, Ikšķile and Salaspils-Laukskola). In all of these places archaeologists have found various man-made artifacts of bone and horn, as well as nests of flint antiquities which suggest that the earliest populations arrived in Latvia at the very end of the late Palaeolithic period (Zagorska, 1992).

This early Latvian residents gradually moved across the eastern part of the country and then moved northward into the Lubāns lowlands, the Vidzeme region and Estonia (Jaanits, 1990). As was mentioned previously, however, the first residents of Lithuania arrived considerably earlier. Moreover, Lithuania was directly affected by several population migrations from the West in the late Palaeolithic period.

Specialists feel that because of these circumstances, two different cultural regions were established in the Baltic during the early Mesolithic period. One conformed to the Niemen Mesolithic culture of Lithuania, while the other involved the Kundian culture in Latvia and Estonia. This cultural division persisted throughout the Stone Age in the Baltic region (Denisova, 1994). Several generations of archaeologists have worked in specifying the genesis of the Kundian culture. Currently specialists are working with a completely new range of archaeological materials from the very last part of the late Palaeolithic period (Zagorska, 1974) and from the early Mesolithic period (Jaanits, 1990). The pre-Boreal period is represented through settlements in Estonia (Pulli, Lepakose) and Latvia (Zvejnieki II) (Zagorska, 1981; 1992; Jaanits, 1990). For the first time, archaeologists have also found well-preserved human skeletons (the Zvejnieki burial ground in northern Latvia) which are dated to the Boreal and Atlantic period (Denisova, 1994; 1996). These data allow specialists to gain a much better understanding of the genesis of the Kundian culture.

The flint industry at the Pulli, Lepakose and Zvejnieki II settlements bore distinct characteristics of post-Svidrian culture (Zagorskis, Zagorska, 1977; Jaanits, 1990). The flint artifacts which have been found at these settlements, moreover, are quite similar to artifacts which have been obtained in late Palaeolithic and early Mesolithic settlements in northern Belarus. These facts have justifiably allowed Dr. K. Jaanits to describe the Krumpleva settlement as a monument to the early phase of the development of Kundian culture and the Daugava river as an important transportation route via which the most ancient populations arrived in Latvia and Estonia (Jaanits, 1990). Searching for indications of Kundian culture in the late Palaeolithic period, Dr. Jaanits has pointed to two possible conclusions. First of all, distinct post-Svidrian traditions in the Kundian culture can be interpreted as pointing to a direct genetic link with Svidrian culture in Poland. Secondly, characteristics of the Svidrian culture flint artifacts are typical of the late Palaeolithic period across a fairly vast section of Eastern Europe (including the Dnieper-Don-Desna, the Volga and the Oka regions). This allows specialists to see roots of the Kundian culture in the late Palaeolithic period in Eastern Europe (Jaanits, 1990).

Our main source of information about the residents of the Mesolithic period and the early Neolithic period in Latvia is the Zvejnieki burial ground. Approximately 300 Stone Age graves were discovered (Denisova, 1975; Zagorskis, 1987). The Zvejnieki burial ground is located in northern Latvia, on the northern shore of the Burtnieks lake and approximately 100 kilometers from the early Mesolithic Pulli settlement of Estonia. Along with the burial ground, the Zvejnieki site also boasts Mesolithic and Neolithic settlements (Zvejnieki II and Zvejnieki I, respectively). The most ancient burial sites in the Zvejnieki area can be dated to the end of the Boreal and the beginning of the Atlantic period (6300 - 5800 BC), and archaeologists have found several individual graves. These graves yielded the most ancient skulls which have ever been found in the Baltic region (Denisova, 1994; 1996). The people who left individual graves here cannot be linked genetically, because ^{14}C dating indicates that the age of the various graves

differs by several hundred years. This means that each person buried at the site can be evaluated separately. In approximately 5000 BC, people at the Zvejnieki site began to bury their dead systematically, and over the course of some 200 years, a late Mesolithic graveyard was established (Denisova, 1996).

During the Atlantic period, there were two culminations of warmer weather which were separated by a brief period of considerably cooler weather. In the northwestern part of Russia, this cooler period lasted from 6200 to 6000 BC (Dolukhanov *et al.,* 1989). In Latvia, the first warmer period took place from 6600 to 6400 BC and coincided with the transfer from the Mesolithic to the Neolithic period. In the Eastern Baltic region during this time, there were several early Neolithic permanent settlements (Zagorskis, 1967; Loze, 1988). During the early Neolithic period in the Baltic region, inhabitants began to fashion dishes out of clay. During the late Mesolithic period, there was a fairly lengthy suspension of new burials in the Zvejnieki burial ground. This suspension began around 4800 BC, and the appearance of new graves in the burial ground bean only after a fairly lengthy pause. Sometime between 4500 and 4400 BC, a new and compact group of graves appeared here. Burial traditions which were unlike those of the late Mesolithic period also appeared (Denisova, 1994; 1996). Judging from anthropological data, there was no direct continuity between those communities which left their graves at this location around the mid-5th century BC and the late Mesolithic populations; distinct anthropological differences existed. This suggests that around the mid-5th century BC there was a new migration of people into Latvia, people who were characterized by the metisized anthropological type. An analysis of anthropological elements in these inhabitants points to distinctly eastern components. Even though no ceramics have been found in the Zvejnieki burial grounds (mid-5th century BC), there is no reason to doubt that the burial grounds belonged to the early Neolithic period. Evidence of this is given by other early Neolithic graves in the Zvejnieki burial grounds (4500-3000 BC), where there was also no tradition of placing clay pots in people's graves.

Currently available data indicate that the most ancient ceramics in the Baltic region can be found in Latvia. In Neolithic settlements, clay ware has become a very typical discovery. The manufacturing of clay vessels in Lithuania began considerably later – at the beginning of the 4th century BC. It appears that clay ware came to Scandinavia around the same time. In the territory around the Baltic region, the most ancient clay vessels have been found in the region which is bordered by the upper Lovate and Daugava rivers. Early Neolithic inhabitants arrived here in the early 5th century BC. It seems that it may well have been from this territory that skills in manufacturing clay vessels may have come specifically from this territory.

To sum up, new facts which have been obtained as the result of archaeological research of the early Mesolithic period in Estonia and Latvia, as well as anthropological data about the most ancient inhabitants of Latvia have allowed specialists to analyze the issue of the most ancient inhabitants of Latvia in a considerably broader context than has been the case until now. Unquestionably, analogous research in neighbouring territories has helped to specify the linked events in the late Palaeolithic period which later affected the most ancient population of the Latvian territory. It appears that we can speak with some certainty of two major directions of ancient migration which during the change from the late Palaeolithic to the Mesolithic period reached the Upper Daugava. Moving along the river, these ancient people came into Latvia. One of these migration directions moved along the southern shore of the North Sea and the Baltic Sea to reach the Pripet basin and the upper Dnieper and Daugava rivers. Movement of populations occurred a number of times over the course of several millennia – in the late Palaeolithic period and in the early Mesolithic period. The second migration wave, it seems, involved the Don river basin, the site of lasting late Palaeolithic, Eastern European settlements. As the result of climactic changes, inhabitants moved northwest along the Dnieper river at the end of the late Palaeolithic period. This direction of movement, which appears only at the end of the late Palaeolithic period and the early Mesolithic period, almost certainly appeared in the early Neolithic and Aeneolithic period, too. It is possible that even in the Stone Age this direction of movement, the main phase of which involved the Dnieper river but which later also reached the

Daugava, played an important role in establishing a communications network between the Baltic and the external world, both to the East and the West (Denisova, 1997).

- **Indo-Europeans in the Eastern Baltic in the view of an archaeologist**

The prehistoric appearance of Indo-Europeans in the Eastern Baltic region is generally thought to be linked with the establishment of the first pre-Baltic and early Baltic territory. This process has been studied both on the bases of archaeological materials and on the basis of the territory in which Baltic hydronyms are found. There has been a search for specific components that may have contributed to the establishment of Indo-European linguistic groups, archaeological or other cultural groups, or various anthropological types. A key element in the Indo-Europeanization of the Eastern Baltic, as well as territories to its East and Southeast, was definitely the appearance of Corded Ware culture, an element which has still not lost its significance in this region.
One hypothesis about the arrival of Indo-Europeans in the Eastern Baltic, Belarus and Central Russia has been proposed by Marija Gimbutas (1963), who has pointed to the Baltic Littoral Piemare culture as the nucleus of Western Baltic culture while linking the beginnings of Eastern Balts with the Middle Dnieper, Fatjanovo and Balanovo cultural groups. She also accepts the idea that bearers of Globular Amphora culture participated in this process. The Indo-Europeanization of the Eastern Baltic has been seen as part of a larger process of Indo-Europeanization in Europe. Links have been drawn with the three-phase expansion of bearers of the Kurgan culture from the steppes of the Black Sea shores and the forested steppes of Ukraine toward Middle Europe (between 4400 and 2800 BC). From there, the second cradle of Indo-Europeans, these people saw the transformation of Globural Amphora, Corded Ware and Baden-Vuchedol cultures and continued their movement to the South and the North, as well as to the Northeast. This hypothesis is based on the identification of a specific economic regime, point out that the bearers of Indo-Europeanization were semi-nomadic and pastoral in their survival strategy. A different economic model is at the basis of a second hypothesis about Indo-Europeanization and its genesis 6,500 years before Christ (there have been calibrated datings of radioactive carbon), when the earliest representatives of land cultivation survival strategies began to migrate gradually from Asia Minor (Anatolia) through the Balkans toward Middle Europe and then on toward the North.
Another hypothesis has emerged more recently and involves placing a center of Neolithization in the region of the Dnieper rapids, utilizing archaeological and genetic arguments and specifying the biological transformation of populations in the region well before any Anatolia impact in the Balkans.
These and other issues of Indo-Europeanization, as well as their link with Neolithization, are fairly important in contemporary scientific literature. In this connection, the first issue that must be reviewed is economic changes in the region which occurred between 3300 and 2000 BC, in order to be able to evaluate the role of these changes in the process of Indo-Europeanization.
A reconstruction of the vocabulary used by Proto-Indo-Europeans shows that they were herders and farmers who were familiar with the plow, who raised wheat and flax, and who were familiar with horses, bulls, pigs and sheep. We can specify and model the process whereby agriculture came to the Eastern Baltic by accepting that the process was peaceful and grounded in impulses gained from farmers in southwestern territories.
The Lubāna lake basin in Eastern Latvia, as well as the Šventoji lagoon and Nida on the Lithuanian shoreline are the main regions where early agricultural efforts were made. These processes must be linked closely to changes in the social structure, ideology and language of society at that time. The Neolithic cultures of these regions, which during the process of Neolithization were representatives of the Post-Narva, Globular Amphora and Corded Ware cultures, drawing some influence from the Funnel Beaker culture; it is from their cultural environment that specialists have been able to draw conclusions about changes which occurred in that period. If we accept that agriculture initially arrived in the Lubāna lake basin as the result

of diffusion with local Post-Narva tribes acquiring skills from the Funnel Beaker culture, then we must give a different interpretation to the early agricultural activities of settlements in the Šventoji lagoon and Nida area. In that case these activities must be linked to the activities of bearers of the Globular Amphora culture who brought new forms of agriculture to the people of the Post-Narva culture.

Archaeological attributes (specific flint-thin butted axes, wooden hoes, earth looseners made of animal horns, flint sickle-shaped knives, grindstones and stone mortars and pestles), as well as the presence of wheat and hemp pollen in the spore-pollen spectra at a region in the Lubāna lake basin where Neolithic settlements were particularly concentrated (the Zvidze settlement), all suggest that land cultivation efforts were first implemented in the region ca. 3300-3000 BC. The discovery of wooden plows, stone hoes and other agricultural tools, as well as seeds from Italian millet two-grain wheat and hemp, meanwhile, suggests that early land cultivation was a fundamental branch of economic activity along the Lithuanian shoreline at a later time (ca. 3030-2580 BC).

The aforementioned Funnel Beaker culture (which spread across a territory that reached from the Netherlands to Poland and from Germany to Denmark and Sweden) and the later Globular Amphora culture, which covered the northern part of Middle Europe, were both cultures in which people maintained an agrarian or an agrarian pastoral economic system.

There are various views with respect to the economic structure of the Corded Ware cultures which participated in the Neolithization of the Eastern Baltic. Some authors have accepted that in Middle Europe, these people were land cultivators, but others deny this, suggesting instead that these were nomadic and pastoral peoples. In order to gain a better idea of the role of this specific culture in the Neolithization process, let us first point to the changes which this culture brought into the cultural environment of Neolithic residents of the Eastern Baltic. Archaeologists more recently have made discoveries, among them grindstones used to grind cereal seeds (both upper and lower stones were found at Abora, in the Lubāna lake basin), stone mortars and pestles, and manual grindstones (large lower stones). This suggests that during the time between 3000/2897 - 2300/2100 BC, initial land cultivation skills came into the middle part of the Eastern Baltic region. It is possible that this process was stimulated by the arrival of early bearers of the Corded Ware culture into the cultural territory of the Post-Narva culture in the Lubāna lake basin. Here, as in the territory of the Piemare culture in coastal Lithuania, they established a local cultural group with new elements, especially in terms of a special sort of clay ware that has become known as corded ware. However, the Corded Ware culture preserved earlier types of tools, including flint arrows and spear tips, as well as stone battle axes of the shafthole type. It can be assumed that it was in this period (around 3000-2500 BC) that early agricultural processes began to stabilize gradually. However, the hemp, wheat and oats, as well as millet and flax, which were grown by the later Piemare culture in coastal Lithuania (along the shore of the Šventoji lagoon and in Nida) suggest that this economic sector was intensified in the period between ca 2400 and 1960 BC.

These facts, which have been learned in the last several years, indicate that the Eastern Baltic became part of the early agricultural zone much sooner than had been assumed previously, as much as 800 years earlier. The link with specific Neolithic cultures allows specialists to view this process on a common Neolithization background, marking the gradual nature of the process, as well as the order in which grain cultures were introduced. In other words, if the Proto-Indo-Europeans really engaged in land cultivation, then the appearance of the Proto-Indo-Europeans in the Eastern Baltic could not have happened earlier than the aforementioned dates.

Changes in language

The Indo-Europeanization of the Eastern Baltic region cannot be seen in full only on the basis of archaeological materials. A review of the cultural groups in the region must also be done.

The appearance of a new type of economy, i.e., of Neolithization, in the westernmost reaches of the forest zone of Eastern Europe, including the Eastern Baltic, and all of the consequences which this process brought in terms of ideology and social structure, may all be linked with the

spread of Indo-European languages. This is all the more likely because the process involved the very earliest communities of farmers and livestock herders, who established contacts among themselves or who were late arrivals in the local Eastern Baltic cultural environment, bringing with them gradual changes in language.

If we accept that the implementation of agriculture in the area occurred gradually, that it started in the eastern part of Latvia (the Lubāna lake basin) between 3300 and 3000 BC, and that somewhat later (3030-2580 BC) it came to the Šventoji lagoon of coastal Lithuania, albeit with a different social structure, including wooden hand plows and models of ox carts, then we can conclude that Indo-Europeanization may also have occurred in a gradual fashion. It brought technical innovation in a variety of areas, and this allowed local communities to expand their vocabularies, either changing existing terminology or accepting completely new words.

Most recently specialists have moved away from the divergence of language in the traditional understanding of the term and have instead begun to look at the convergence of language, i.e., the gradual change in language which occurs when outside contacts are made. This approach may offer greater opportunities to interpret the process in the archaeological sphere: there have already been attempts to link the spread of Indo-European languages directly with the arrival of agriculture and livestock herding, as had previously happened among hunters and gatherers who established contacts with Neolithic culture representatives who had an agricultural or an agricultural and pastoral lifestyle. This thesis has been extensively substantiated and has been accepted in previous research. It must be pointed out, however, that this is only one of a number of possible routes by which we can study the Indo-European issue. Research of the implementation of agriculture and livestock tending in the Eastern European forest zone has become active only recently, and along with innovations in the social and ideological sphere, we may also discover new opportunities to model changes in language.

The fact that the center of Neolithization moved to the Dnieper rapids region means that we must devote far more attention to the Dnieper river than has been done until now. The fact that regions of the Middle Dnieper and the Upper Dnieper were subjected to processes of Indo-Europeanization has been discussed in the literature extensively, but the question remains whether the process perhaps did not occur only by way of central Europe, instead coming directly up the Dnieper river. Given that linguistic and economic changes can be viewed together (this has been accepted in the Indo-Europeanization models proposed by M. Gimbutas and C. Renfrew) we must note that the Indo-Europeanization of the Eastern Baltic was not concluded by representatives of Globular Amphora and Corded Ware culture who arrived in the region. Their successors, who were representatives of hybrid cultures from the Piemare region in coastal Lithuania and the Lubāna lake basin (people who had a specific method of manufacturing and ornamentation corded ware), continued the process through the Aeneolithic period, as well as the early Bronze Age, because the permanent implementation of the new economic branch could only have happened in that period. In that case the process may also have involved other cultures from the period, including the Northern Belarussian culture which has been found at some settlements near the Lubāna lake basin, as well as the Marjanov culture from the Middle Dnieper. These cultures appeared in the archaeological map of Eastern Europe quite recently, and chronologically they correspond to the period of the Bell Beaker culture in Central and Western Europe. Perhaps it is this period in the forest zone of Eastern Europe which contains new possibilities in resolving the Indo-European question. We must not, however, forget that in addition to the ancient Indo-Europeans, the territory in question also hosted Finno-Ugric cultures and their successors.

There is some evidence that a new cultural unit was established in the Eastern Baltic forest zone in the early period of the Bronze Age. Specialists have pointed to the appearance of ceramic ware with signs of the Fatjanovo culture in late Volosovo culture settlements between the Volga and the Oka basin (more recently this has been deemed an example of Saghar ceramic ware), and to the fact that a special Saghar culture was established in the Middle Oka basin which had technological and ornamental methods reminiscent of the Fatjanovo Corded Ware culture. It is

also true that these elements spread to the Lubāna lake basin, to settlements in Northern Belarus, and to the left shore of the Polesje (the Desna, Seima and Soža basin), as well as its right shore (the Pripet basin). These cultures established a cultural and historical situation which quite possibly provided a greater or lesser framework for the further spread of Indo-Europeanization (Loze, 1997).

3. AIMS OF THE PRESENT STUDY

The main goal of the present study was to obtain the knowledge about mtDNA variation in Latvia and its sub-populations, and further to perform a large-scale phylogeographic analysis at inter- and intra-population level in order to understand the origin and rise of the genetic diversity of Latvian mtDNA lineages. In such a way the observed matrilineal genetic composition may shed the light on the numerous questions of Latvian genetic history. In order to draw conclusions about the process that might have been involved in the shaping of present mtDNA diversity in Latvians, the aim of this study was to examine mtDNA variability in comparison with surrounding Indo-European and Finno-Ugric speaking reference populations.
We have been also interested in Saami, earlier shown to be genetic "outliers" in the European genetic landscape. We tried to clarify the problem of their position among European populations, studying which mtDNA lineages are spread among the Saami in a wider Eurasian context – where did these lineages possibly arise, how did they reach the northernmost Fennoscandia and are the Saami indeed "outliers" among European populations, or simply a small distinct part of the European unity.

The following questions were posed:

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

2. How does the genetic variability of the Latvian mtDNA pool reflect their linguistic background compared to neighboring populations, in particular Finno-Ugric and Slavonic speakers?

3. How does the pattern of mtDNA variation in Latvians correspond to that observed for their Y chromosomes, where the Latvians and Lithuanians, contrary to their linguistic affinity, are very close to their Finno-Ugric-speaking neighbours?

4. From a microevolutionary point of view, do the two extant Baltic-speaking populations – Latvians and Lithuanians – display distinct maternal lineages that can be considered as region-specific for the two populations?

5. What is the phylogenetic affiliation of mtDNA spread among the Saami in the Eurasian context?

4. SUBJECTS AND METHODS

4.1. Subjects

The experimental basis of the current thesis relies on the analysis of the mtDNA variation in the Latvian population as a whole and it subpopulations, analyzed by the author: 351 Latvians representing four anthropologically, archaeologically and ethno-linguistically different regions of Latvia: 88 from North-Western region (Northern Curonia), 67 from the Central region (Semigalia), 68 – from the South-Western region (Southern Curonia), and 128 from Eastern region (Lettigalia) (figure 7).

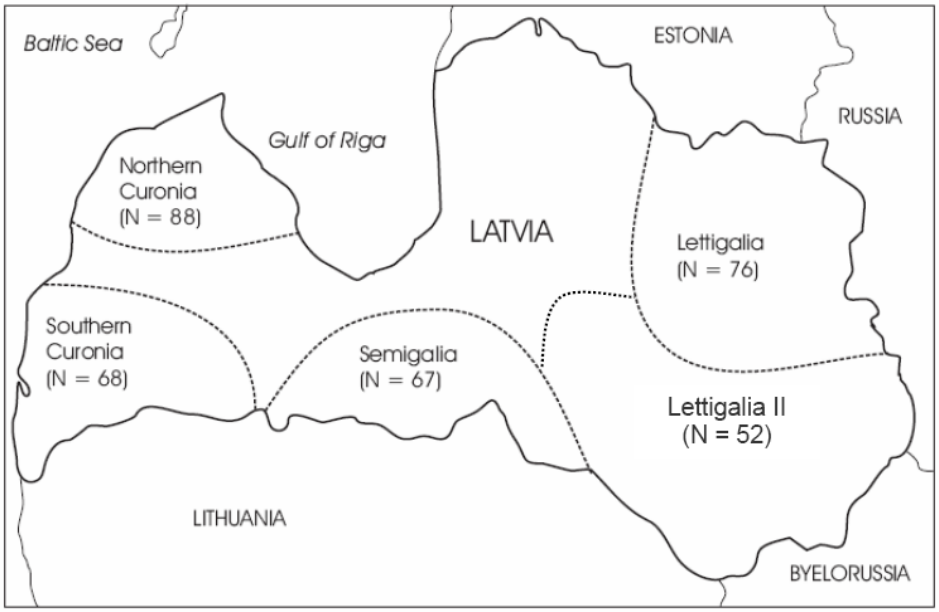


Figure 7. The four Latvian regions from where the samples of the present study have been collected are shown with a dashed line. The Sample size (N) for each region is given in brackets.

Here, it needs to be mentioned that we have expanded our mtDNA dataset of Latvians added to the previously published data (ref. **I**) 52 samples from south-eastern region of Latvia. DNA samples were collected and analyzed in collaboration with Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu (prof. R. Villems); Riga Stradins University (prof. A. Krumina) and the Institute of the Immunology (prof. A. Socnevs). DNA samples were collected from healthy unrelated individuals after obtaining informed consent, and their ethnicity, as well as maternal ancestry over the last three generations, was established from interviews. The Ethics Committee of the Rigas Stradins University approved the research protocol.

For comparison, the mtDNA database of 11,236 individuals, consisting of published mtDNA data (Lahermo *et al.,* 1996; Baasner *et al.,* 1998; Comas *et al.,* 1998; Lutz *et al.,* 1998; Kittles *et al.,* 1999; Pfeiffer *et al.,* 1999; Crespillo *et al.,* 2000; Dimo-Simonin *et al.,* 2000; Helgason *et al.,* 2000, 2001; Pereira *et al.,* 2000; Richards *et al.,* 2000; Cali *et al.,* 2001; Larruga *et al.,* 2001; Malyarchuk & Derenko, 2001; Meinilä *et al.,* 2001; Mogentale-Profizi *et al.,* 2001; Tagliabracci *et al.,* 2001; Bermisheva *et al.,* 2002; Derbeneva *et al.,* 2002a, 2002b; Malyarchuk *et al.,* 2002, 2003; Passarino *et al.,* 2002; Derenko *et al.,* 2003; Dubut *et al.,* 2004; Kasperavičiūtė *et al.,* 2004; Loogväli *et al.,* 2004) as well as unpublished mtDNA data from 4,732 individuals from different Eurasian populations (data base of Department of Evolutionary Biology, prof. R. Villems), were used as background information for the analysis.

4.2. Genotyping of mtDNA variation

DNA extraction and amplification

DNA was extracted from venous blood using the standard phenol–chloroform method as described in Sambrook (1989). DNA fragment, encompassing the mtDNA HVS-I between nps 16024-16383, was amplified and sequenced in all samples using forward (L15926) and reverse (H16498) primers (Vigilant *et al.,* 1989). In addition, mtDNA HVS-II sequences between nps 70–350 (primers from Calloway *et al.*, 2000) were determined only for selected samples from hgs U4, H and W. Amplification reactions were performed on 10 ng of template DNA in a 20-µl volume by use of Ampli*Taq* (*Fermentas*, Lithuania) or FIREPol (*Solis Biodyne*, Estonia) DNA polymerase I. The cycle profile started with 94^0C for 2 min, followed by 35 cycles of 94^0C for 30 sec, 58^0C for 20 sec, 72^0C for 30 sec, and 72^0C for 3 min (UNO II Thermocycler, Biometra, Germany). Negative controls were prepared for both the DNA extraction and the amplification processes.

mtDNA Sequencing and Genotyping

Prior to sequencing the PCR products were purified with shrimp alkaline phosphatase (SAP) and exonuclease I (ExoI) (Werle *et al.,* 1994). PCR products were sequenced directly from both strands by use of the DYEnamic™ ET terminator cycle sequencing kit (Amersham Pharmacia Biotech, Sweden) according to the manufacture's protocol. The cycle-sequencing profile was 32 cycles of 96^0C for 10 sec, 50^0C for 5 sec, and 60^0C for 4 min. Sequencing reactions were run on the MegaBace 1000 DNA (Amersham Pharmacia Biotech, Sweden) and ABI Prism 3100 DNA automated sequencer (Applied Biosystems, USA). The sequences were compared with the revised Cambridge Reference Sequence (rCRS Andrews *et al.,* 1999) by use of the Genetics Computer Group Wisconsin Package or by Contig Express software (Invitrogen, USA). All mutations are in this study reported as differences from rCRS; only transversions are further specified. Length polymorphism of the A and C stretches between nps 310–315 and 16180–16188 were disregarded in the analysis. Similarly, transversions adjacent to the poly-C tract in positions 16184–16193 were ignored as probable sequencing artefacts. Following the guidelines the data were checked for the presence of systematic sequencing errors (Bandelt *et al.,* 2002).
To confirm the haplogroup affiliations of mtDNA sequences, hierarchical RFLP analysis was performed using 17 restriction endonucleases: 73*Alw*44I, 3007*Bsh*1236I, 4332*Eco*47I, 4577*Nla*III, 4643*Rsa*I, 4769*Alu*I, 4793*Bsu*RI, 4831*Hha*I, 5003*Dde*I, 7025*Alu*I, 8249*Ava*II, 8446*Ssp*I, 8994*Hae*III, 10032*Alu*I, 10397*Alu*I, 12308*Hinf*I, 13704*Bst*OI, 14465*Acc*I, 14766*Mse*I, 15606*Alu*I, 15904*Mse*I and 16487*Dde*I. For detection of the mutation at np 456 (from C to T) allele-specific PCR was used (Loogväli *et al.,* 2004). Nucleotide variants at coding region sites 5656, 6776, 7385, 10927 and 11812 were ascertained by sequencing for all U5 samples (Tambets *et al.,* 2004).
Classification of haplogroups (hgs) and sub-haplogroups (sub-hgs) was based on affirmed nomenclature (Torroni *et al.,* 1993, 1994, 1996; Richards *et al.,* 1998; Macaulay *et al.,* 1999a; Finnilä *et al.* 2001; Kivisild *et al.,* 2002; Loogväli *et al.,* 2004; Tambets *et al.,* 2004). Simplified haplogroup detection scheme can be observed in Table 2.

Table 2. Detection scheme of principal mtDNA haplogroups using hierarchical RFLP analysis

HVS-I haplotype (-16 000)	Hg	73 Alw44I	-7025 AluI	-14766 Tru1I	-4577 Hin1II	+12308 HinfI	+4646 RsaI	+15606 AluI	-13704 MvaI	-8994 BsuRI	+10032 AluI	+14465 AccI	+ 10397 AluI
			H	HV	V	U	U4	T	J	W	I	X	M
CRS	H	A	-										
129	H	A	-										
189-356	H	A	-										
153-298	V	A			-								
311	HV	A		-									
356	U4	G				+	+						
189-248-270	U5	G				+							
189-270	U5	G				+							
069-126	J	G							-				
069-126-145-172-222-260-261	J	G							-				
126-294-296-304	T	G						+					
126-163-186-189-294	T	G						+					
129-172-223-311-319-391	I	G									+		
129-172-223-311-355-391	I	G									+		
223-292	W	G								-			
192-223-292-325	W	G								-			
189-223-266-278	X	G										+	
093-223-227-278-362	M	G											+

CRS – Cambridge Reference Sequence

4.3. Data analysis

Haplogroup and haplotype frequencies as well as diversities (calculated as in Nei 1987) of particular clusters and sub-clusters were calculated and compared, including the available published data. Statistical analysis was based on mtDNA haplotypes that were classified into hgs.

The distribution of mtDNA diversity was measured using the analysis of molecular variance (AMOVA, Excoffier *et al.,* 1992) as variation within and between population groups, which were composed either on the basis of linguistic affiliations, according to language subfamilies (Baltic-, Slavonic-, Germanic-, Finno-Ugric-speaking populations), or by geographical location of the studied populations (see ref. **I**). In order to investigate the population structure of Latvians, all four Latvian ethnolinguistic groups were first treated separately and then grouped into two main subgroups – Curonians (Northern and Southern Curonians) and Central/South-Eastern Latvians (Semigalians, Lettigalians) – according to their geographic and linguistic affiliations. The significance of the results was tested by 10,000 permutations. Standard errors were estimated from 1000 bootstrap iterations. For AMOVA analysis and calculation of pairwise genetic distances (*F*st) the ARLEQUIN 3.11 package (Excoffier *et al.,* 2005) was used. Population data were compared by use of principal component analysis (program POPSTR), based on haplogroup frequencies.

For further analysis, phylogenetic networks relating different haplotypes (as described in Bandelt *et al.,* 1995; 1999) were constructed. The statistical significance of population differences with respect to the frequencies of mtDNA hgs was evaluated using the chi-square test (uncorrected for multiple comparisons). The time estimates for founding haplotype clusters were calculated by use of statistic ρ, the average transition distance from the putative founder sequence, and

calibrated using a transition rate of 1 in 20,180 years for the HVS-I region between nps 16090-16365 (Forster *et al.,* 1996). The standard error (σ) was calculated as in Saillard *et al.,* (2000).
Definitions of statistical parameters and methods that were mentioned in this section are given herein briefly and were taken from Jobling, Hurles and Tyler-Smith (2004), if have not been specified in the text.
Analysis of Molecular Variance (AMOVA) is a method of estimating population differentiation directly from molecular data and testing hypotheses about such differentiation. This method takes into account the molecular relationship of alleles, rather than just their frequencies, when apportioning variance between tiers of the hierarchical population structure.
A variety of molecular data – molecular marker data (for example, RFLP or AFLP), direct sequence data, or phylogenetic trees based on such molecular data – may be analyzed using this method.
Permutation is the rearrangement of objects or symbols into distinguishable sequences. Each unique ordering is called *a permutation*. For example, with the numerals one to six, each possible ordering consists of a complete list of the numerals, without repetitions. There are 720 total permutations of these numerals, one of which is: "4, 5, 6, 1, 2, 3".
Bootstrapping is a statistical method for estimating the sampling distribution of an estimator by sampling with replacement from the original sample. It may also be used for constructing hypothesis tests. It is often used as a robust alternative to inference based on parametric assumptions when those assumptions are in doubt, or where parametric inference is impossible or requires very complicated formulas for the calculation of standard errors.
F_{ST} measures the effect of population subdivision, which is the reduction in heterozygosity in a subpopulation due to genetic drift. F_{ST} is the most inclusive measure of population substructure and is most useful for examining the overall genetic divergence among subpopulations. It is also called coancestry coefficient (q) or 'Fixation index' and is defined as correlation of gametes within subpopulations relative to gametes drawn at random from the entire population (**S**ubpopulation within the **T**otal population). It is calculated as using the subpopulation (average) heterozygosity and total population expected heterozygosity. F_{ST} is always positive; it ranges between 0 = panmixis (no subdivision, random mating occurring, no genetic divergence within the population) and 1 = complete isolation (extreme subdivision). F_{ST} values up to 0.05 indicate negligible genetic differentiation whereas >0.25 means very great genetic differentiation within the population analyzed. F_{ST} is usually calculated for different genes, and then averaged across all loci, and all populations. For human populations, the average value of F_{ST} for a large number of DNA polymorphisms is 0.139 (and 0.119 for non-DNA polymorphisms) (Cavalli-Sforza *et al.,* 1994).
Having obtained a set of pairwise genetic distances between a set of populations, we can display this information in a comprehensive manner. It is difficult to detect patterns from a table of pairwise differences; therefore, graphical displays are preferable. If we have *n* populations, we require *n*-1 dimensions to fully display their pairwise genetic distances as graphical distances. Multivariate analyses allow us to reduce multidimensional space to the two or three dimensions we can comprehend, while reducing the inevitable loss of information. **Principal Component Analysis** (PCA) is a commonly used example of this approach. Individual axes, known as principle components (PCs), are extracted sequentially, with each PC encapsulating as much of the remaining variation as possible. Using PCA it is possible to estimate the proportion of the total variance in the total dataset that has been summarized within these reduced dimensions. Figure 8 shows how PCA reveals global relationships from a set of pairwise distances between populations.

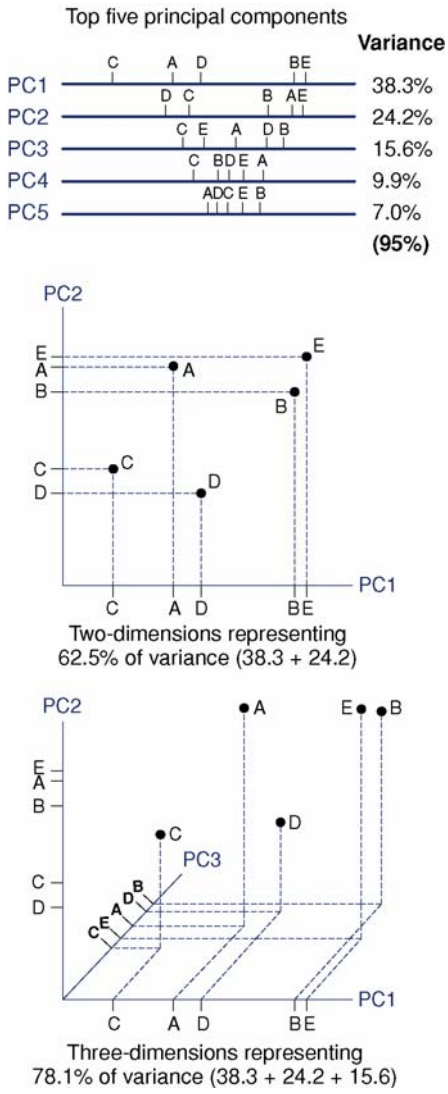


Figure 8. Graphical representations of principal components analysis of five populations in both two and three dimensions (Jobling, Hurles and Tyler-Smith 2004). Principal components are extracted from multivariate data from five populations (A-E) such that each successive PC contains a smaller proportion of the overall variance. Each PC is displayed as a single axis. The first five principal components account for 95% of the variance within the dataset.

In probability theory and statistics, the **chi-square distribution** (also **chi-squared** or **χ^2 distribution**) is one of the theoretical probability distributions most widely used in inferential statistics, i.e. in statistical significance tests. It is useful because, under reasonable assumptions, easily calculated quantities can be proven to have distributions that approximate to the chi-square distribution if the null hypothesis is true.

Dating using the statistic ***ρ* (rho)** requires the construction of a phylogeny relating the intra-allelic diversity. This phylogeny must contain the root haplotype, and so called median networks are often used, as they allow the reconstruction of ancestral haplotypes, even if they are not observed within sample. The ρ statistic represents the average number of mutational changes between the root haplotype and every individual in the sample. These mutational changes are counted from network itself, rather than by estimation from observed number of mutational differences between two haplotypes; this takes account of possible reversions and parallelisms at sites with greater mutation rates. This statistic is simply related to time by the equation: $\rho = \mu t$, where μ is a mutation rate and t – time to MRCA.

Phylogenetic networks and trees were described in section 2.3. in Literature overview.

5. RESULTS AND DISCUSSION

In the list of publications the articles are not ordered according to their publishing time. References **III** and **IV** (review articles that were published in *Proceedings of the Latvian academy of Sciences*) are mostly devoted to the structural organization and worldwide phylogenetic studies of both uniparentally inherited markers, mtDNA and Y chromosome; therefore, they have not been overviewed in the following sections.
There are several possibilities to analyze the data of uniparentally inherited marker systems. One way is to compare the mtDNA variation (*e.g.* frequencies of haplogroups) between different populations or population groups in order to find out the similarities and differences in the composition of their maternal gene pools (reference **I**). Alternatively, one may apply a phylogeographic approach in order to study the global spread of particular lineages (*e.g.* Torroni *et al.,* 2001b for mtDNA hgs V and pre-V; Reidla *et al.,* 2003 for mtDNA hg X). The two approaches can be combined (reference **II**), so that a wide phylogeographical analysis of the major maternal lineages present in a population (here, in the Fennoscandian Saami) can provide information, which helps to shed light on the pre-history of the population under study. This approach can be particularly informative in combination with a parallel and phylogeographically comparable study of Y-chromosomal heritage. In this dissertation, the main stress will be on the results associated with mtDNA analysis. In addition, we have performed micro-phylogeographical analysis of maternal lineages that are spread in different ethnolinguistic groups of Latvia.

5.1. The phylogenetic affiliation of the maternal lineages of the Latvian population on the European mtDNA tree (Ref. I)

The initial purpose of our study was to investigate so far poorly described Latvian population, to obtain data of sufficient phylogenetic resolution and to apply the phylogeographic approach in order to study the spread of mtDNA lineages in Latvia in a comprehensive context to their distribution in different European populations. However, sub-populations of Latvia were not surveyed in details in reference **I**, therefore, section 5.1.1. is mostly dedicated to mtDNA haplogroups and lineages distribution in their.

5.1.1. Regional differences among the Latvians: a pattern of mtDNA lineages in four ethnolinguistic sub-populations of Latvia

The observed mtDNA haplogroup and sub-haplogroup frequencies for the Latvian sub-populations are presented in table 3. Most of them fall into the variability of European populations and cannot differentiate the studied samples (although appearing quite heterogeneous). Additional two haplogroups have been found among Latvian mtDNA pool (U8 and U*) at negligible frequencies, which are also spread at low frequencies in different southern and north-eastern European populations (see table 2, ref. **I**).
However, peculiar frequency distributions across our Latvian sample can be noted. Statistically significant differences ($p \leq 0.05$) in haplogroup frequencies between the eastern and the western/central parts were seen within four sub-lineages of one of the most diverse and abundant haplogroup among Latvians, U (U2, U4, U5 and K, correspondingly). 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, which is close to the frequency of occurrence of U4 observed in the Volga-Uralic region (Bermisheva *et al.,* 2002). Perhaps more importantly, the heterogeneity of hg U4 is also relatively high in Semigalia: out of thirteen U4 haplotypes found among Latvians eight were observed there. Besides, the frequency of U4 was significantly higher in Semigalia than in Lettigalia ($p \leq 0.05$).

Table 3. mtDNA haplogroup frequencies (%) among Latvian sub-populations

Haplogroup	North Curonians (NC)		Semigalians (SE)		South Curonians (SC)		Lettigalians (LE)		Latvians (total)	
	n	%	n	%	n	%	n	%	n	%
H	48	54,5	31	46,3	26	38,2	47	36,7	152	43,3
HV	1	1,1	1	1,5	2	2,9	3	2,3	7	2,0
V	3	3,4	2	3,0	1	1,5	4	3,1	10	2,8
J	6	6,8	4	6,0	7	10,3	6	4,7	23	6,6
T	10	11,4	3	4,5	9	13,2	11	8,6	33	9,4
U*	0	0,0	0	0,0	0	0,0	1	0,8	1	0,3
U2	0	0,0	2	3,0	0	0,0	7	5,5	9	2,6
U3	0	0,0	2	3,0	2	2,9	2	1,6	6	1,7
U4	6	6,8	10	14,9	7	10,3	6	4,7	29	8,3
U5a	4	4,5	6	9,0	5	7,4	13	10,2	28	8,0
U5b	0	0,0	2	3,0	2	2,9	9	7,0	13	3,7
U8	0	0,0	0	0,0	0	0,0	1	0,8	1	0,3
K	0	0,0	1	1,5	1	1,5	6	4,7	8	2,3
W	5	5,7	1	1,5	3	4,4	3	2,3	12	3,4
X	0	0,0	0	0,0	0	0,0	1	0,8	1	0,3
I	5	5,7	2	3,0	3	4,4	7	5,5	17	4,8
M (G)	0	0,0	0	0,0	0	0,0	1	0,8	1	0,3
Sample size	88		67		68		128		351	
Gene diversity	0.9470 ± 0.0148		0.9815 ± 0.0081		0.9750 ± 0.0101		0.9822 ± 0.0039			

The remainder examples of region-specific differences in the Latvian mtDNA pool are observed mostly between Western (Northen Curonia) and Eastern (Lettigalia) parts of Latvia and are provided by sub-hgs U2, U5b, and K. Sub-hg U2, which had a particularly high frequency in Eastern Latvia – seven out of nine Latvian sub-hg U2 mtDNAs were found in the Lettigalia. The frequency of U2 was significantly higher in Lettigalia than in the Western parts of Latvia– Northern ($p \leq 0.01$) and Southern Curonia ($p \leq 0.025$). The high frequency of hg U2 in Lettigalia could be best explained by a recent founder effect, because all U2 genomes found belong to the same HVS-I haplotype (see table 1, ref. **I**). Sub-hg U5b, which again appears more prevalent in Eastern Latvia – nine out of thirteen Latvian sub-hg U5b mtDNAs were found in Lettigalia. The distribution of U5b mtDNAs was significantly higher in Lettigalia than in Northern Curonia ($p \leq 0.025$), where sub-hg U5b is virtually absent (see also section 5.2.2.). The other hg that has a different frequency pattern in Eastern and Western Latvian sub-populations (table 3) is K; statistically significant differences of hg K frequencies were observed between Northern Curonia and Lettigalia, with hg K being more common among the latter ($p \leq 0.05$). All hg K mtDNAs from Latvians shared one HVS-I-haplotype with a 16224–16311 motif. However, according to Finnilä *et al.,* (2001), this motif may be characteristic of many hg K branches, which differ at several coding region nps and have therefore been phylogenetically separated for a long time.

The majority of unique sequences found in Latvians nevertheless derive from the HVS-I sequence variants that are common in most European populations. For example, haplotype 16080–16129–16142–16189–16356 from sub-hg H1 was found in Central and Western regions of Latvia, but has not been reported in published datasets. However, it is a two-step derivative of the HVS-I motif 16080–16189–16356 that is widely spread in Europe (see also figure 3, ref. **I**). Meanwhile, we found four unique representatives of sub-hg H7 in our sample of Latvian mtDNAs (see table 3, figure 3, ref. **I**). All of them shared the HVS-I motif 16217–16311 and were collected from three different Latvian regions (Eastern, Central and Western Latvia), which have not been described previously. From Latvian mtDNA pool we detected only a single member of Asian-specific hg M – G2a – in Lettigalia, the Eastern part of Latvia.

To get a more detailed picture of matrilineal genetic composition of the Latvian sub-populations we have performed several statistical approaches: calculation of pairwise genetic distances (*F*st), PC analysis and AMOVA. Pairwise *F*st distances were calculated using two different

classifications in order to determine the closest Latvian sub-populations: 1) based solely on the haplogroup frequencies; 2) based on the frequencies of all haplogroups and HVS-I haplotypes combined (table 4). One can observe that sub-populations of Northern Curonia and Lettigalia produced *p* values less than 0.01, moreover, after applying the second classification, it appears that statistically significant differences ($p \leq 0.05$) were observed for three Latvian sub-populations, obviously, Northern Curonia and Lettigalia, and more intriguing, Semigalia. Albeit, this excess of low *p* values emerged from its regional differences can be attributed to the outlying frequencies of haplogroups U4, U5b and U2 in Central and Eastern regions of Latvia (Semigalia and Lettigalia) (table 3). The analysis indicates that genetic distances between sub-populations can be observed already at haplogroup level, however, additional information from HVS-I haplotypes has provided a better resolution in attempt to elucidate the genetic relationship between different ethnolinguistic groups of the Latvian population.

Table 4. Pairwise *F*st distances between Latvian sub-populations (A – based on the haplogroup frequencies; B – based on the frequencies of all haplogroups and HVS-I haplotypes combined)

A.

	NC	SC	SE	LE
NC				
SC	0.00933			
SE	0.00481	0.00000		
LE	0.02141**	0.00000	0.00477	

B.

	NC	SC	SE	LE
NC				
SC	0.00066			
SE	0.00655*	0.00000		
LE	0.00732**	0.00000	0.00000	

**$p<0.01$; *$p<0.05$; others > 0.05. Population codes are given in alphabetical order as follows: LE – Lettigalians; NC – Northern Curonians; SC – Southern Curonians; SE – Semigalians.

A principal component (PC) analysis distinguished Lettigalians from the other Latvian sub-populations, in particular, from Northern Curonians. The difference revealed by the first component, which accounts for 61% of total variation, is likely due to the presence of sub-hgs U2, U5b and K at relatively higher frequencies in Lettigalia than in other Latvian sub-populations clustering them into eastern and western affiliation. The second component (22%) reflects mainly a different frequency distribution of hg U4, which is at a higher proportion in Central region of Latvia (Semigalia) than in Western and Eastern Latvian sub-populations. Sub-population from Southern Curonia occupies the intermediate positions between Northern Curonians, Semigalians and Lettigalians (Figure 9).

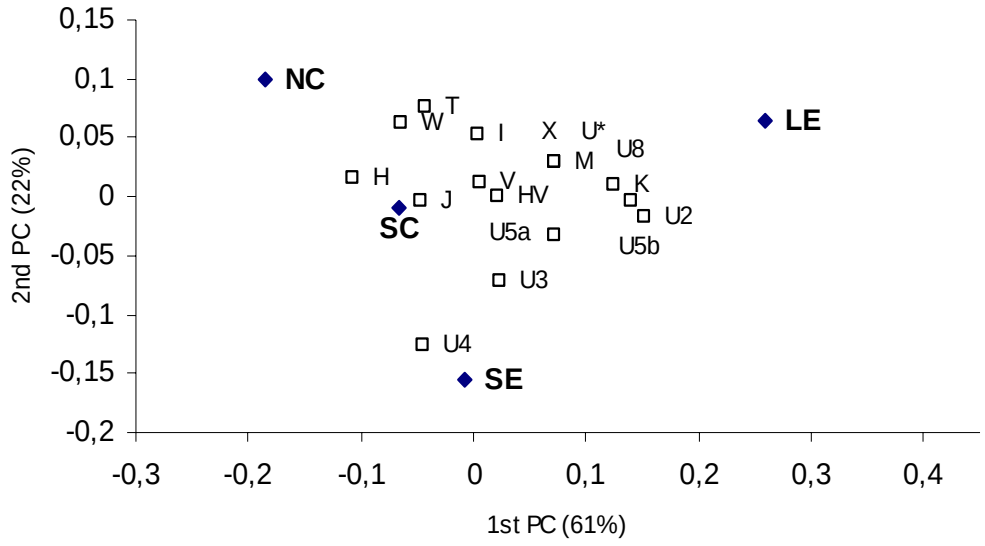


Figure 9. Principal component (PC) analysis based on mtDNA haplogroup frequencies of Latvian sub-populations. Population codes are given in alphabetical order as follows: LE – Lettigalians; NC – Northern Curonians; SC – Southern Curonians; SE – Semigalians. The genetic variation retained by different components is shown in brackets.

By using AMOVA, we estimated different components of the observed genetic variance (see table 4, groupings F and G, ref. **I**). Almost all (~99.5%) intra-population variation fell within ethnolinguistic groups. However, it has to be noted that combination G produces the highest variation among groups (0.49%) and the lowest variation within groups (0.05%), associated with strong significance for the latter ($p \leq 0.029$). In reference **I** we have not performed the comprehensive intra-population analysis of the Latvian sub-populations; therefore, we were not able to distinguish phylogeographic structure of mtDNA variation within ethnolinguistic groups. It needs to be mentioned that intra-population homogeneity was observed among Lithuanians (Kasperavičiūtė *et al.,* 2004). Thus it is likely that even if genetic differences between Baltic tribes existed in the past, they disappeared during the last millennium. However, the sample size (30 individuals) used in this study might be too small to exclude the existence of minor differences between the ethnolinguistic groups of the Lithuanian population.

Frequency distribution of different haplogroups and heterogeneity of mtDNA lineages had similar bipolar structure, thus reflecting a division into eastern and western sup-populations of Latvia. This is supported also by the *F*st values that suggest clear differences in mtDNA variation between the geographical groups even within the most common haplogroups. Furthermore, the proportion of variance between eastern/central and western groups of 0.49% in the AMOVA analysis can be considered to represent substantial genetic differentiation in a culturally and historically heterogeneous country. $\chi 2$-test, AMOVA and PCA analyses results yield interesting information on the internal history of the Latvian population and its sub-populations. The genetic trace of the forced internal migration movement from eastern/south-eastern region (Lettigalia) towards the north-western and central parts of the country can be seen based on mtDNA variation, and is in accordance with archaeological data.

The first residents in Latvia arrived not through Lithuania, as would seem logical, but rather from the Southeast, using the Dnieper River and the Upper Daugava for this purpose. This early Latvian residents gradually moved across the eastern part of the country and then moved northward into the Lubāns lowlands, the Vidzeme region and Estonia (Jaanits, 1990). The work of anthropologist Raisa Denisova has been of particular importance in learning about the ethnic identity of people who lived in eastern Latvia before the Lettigallians and Selonians. Denisova has found craniological materials in 17 burial grounds, as well as individual graves, which unquestionably points to the fact that before the Lettigallians and Selonians arrived in eastern Latvia from the South, the area was populated mostly by the Baltic's Eastern Finns who, along with the residents of south-eastern and north-eastern Estonia, as well as north-western Novgorod Major, made up a unified group, at least from an anthropological perspective. Ethnic processes in Latvia during the early metals age was a direct continuation of those processes which took place in the territory during earlier periods of inhabitancy, especially those which occurred during the middle and late Neolithic period. 1) In the Eastern Baltic, the establishment of Latvia's two indigenous ethnic groups – the Balts and the Finns – occurred under the direct influence of the extensive migratory processes which took place at that time. 2) There is reason to believe that the consolidation of these ethnic groups began with the appearance of the pit-comb pottery culture in the middle of the third millennium BC. As a result of this process, two anthropological types of humans appeared: the European type, which existed since the Mesolithic period, and the Metis type, which was created when Mongoloid characteristics were stratified upon the European type. The first group maintained genetic contacts to the South-Southeast, while the second group had genetic links to the East. This was of significance in subsequent ethnic processes, as well. 3) Ethnic processes became more complex at the turn of the third millennium BC, when corded ware culture bearers entered the Eastern Baltic, the Southwest of Finland, and the forested zones to the East. These new arrivals were land cultivators and livestock tenders, and specialists consider them to be "primary Balts", and as they flowed into an ethnically heterogeneous environment where fishermen and hunters lived, complex ethnic processes ensued. As the result of these, Finnish ethnic groups consolidated in the northern sections of the Baltic territory, and

Baltic ethnic groups began to dominate the southern sections. The fact that this occurred at the end of the second millennium BC and during the first millennium BC is evidenced by archaeological monuments typical of the two groups. 4) The border zone between Finnish and Baltic lands crossed the middle part of Latvia in the first millennium BC – along the Venta and Abava rivers in Courland, along the southern edge of the Vidzeme highlands near Saulkrasti, and then toward the Lubāna lake in Latgale. During the early Iron Age, there were several large regions in what is now Latvia where populations were linked more closely by various economic, personal, religious and, one can assume, administrative and political contacts. Strengthening of these links facilitated the cultural heterogeneity of various regions within the territory, something that presumably was also caused by ethnic differences. These differences were manifested in burial traditions, ornamentation, and occasionally in tools and pottery. It should be noted that none of these regions lay entirely within Latvia – each of them also covered some territory in neighbouring countries. The northernmost border of the Latvian territory which was occupied by the ancestors of the Curonians cannot be specified with any precision. Judging from the distribution of gravesites, it probably did not stretch much beyond the Durbe lake. The discovery of a few scattered sites allows specialists to move the border a bit further to the North, perhaps to Kazdanga. Even farther to the North – up to the Abava river and the Lower Venta, there is a broad zone in which archaeologists have found no traces of the early Iron Age. Apparently this was a sparsely populated region which separated Balts to the South from Baltic Finns to the North. In the territory which was populated by the ancestors of Curonians, no paleoanthropological materials from the early Iron Age have been discovered, so there is no possibility to provide a physical characterization of the inhabitants from that time. Based on various indirect evidence one can conclude that the ancestors of the Curonians had a gracile and narrow-faced anthropological type (Vasks, 1997). Thus the contemporary population of Latvia, the subject of the present study, is composed of a complex mixture of former Baltic tribes, with potentially varying influences from Finno-Ugric and Slavic sources, which could have resulted in genetic heterogeneity within Latvia.

Our detailed study confirms the existence of population substructure within a population that has earlier been considered homogenous based on mtDNA data (ref. **I**). Latvia is not the first population where such conclusions have proven to be false, as the Icelandic and Finnish populations have also been shown to have significant substructuring (Helgason *et al.,* 2005; Lappalainen *et al.,* 2006). These could be seen as warning examples regarding *a priori* assumptions of internal homogeneity for small and culturally non-uniform populations.

Undetected population structure is a well-known potential source of error in case-control association studies, especially if the sampling strategy is different for cases and controls in terms of geographical origin. In such cases, differences in allele frequencies between cases and controls may be due to geographical stratification of allele frequencies instead of an association to a disease, which can thus lead to false positive and negative results (Freedman *et al.,* 2004; Marchini *et al.,* 2004). The sensitivity of mtDNA to the population substructure makes it a good marker for detecting regional variation, but similar patterns are not necessarily present in Y-chromosomal and autosomal variation.

5.1.2. Review of Latvian genetic heritage as revealed by mitochondrial DNA lineages (ref. I)

mtDNA variation was investigated in a sample of 351 Latvians, a Baltic-speaking population from Eastern Europe. Sequencing of the HVS-I in combination with analysis of informative coding region markers revealed that the vast majority of observed mtDNAs belong to haplogroups common to most European populations. By incorporating 52 samples from Latvian eastern region (Lettigalia) to the previously published dataset of Latvians (299 samples) we have not observed pronounced differences in haplogroup distribution. Nevertheless, the proportion of hgs U4 and U5 became more biased to the latter one (8.3% and 11.7%, correspondingly, table 3), and it is due to the presence of sub-hg U5b at higher frequency in Lettigalia. Yet it is worthwhile

to note here that the uniformity of the genetic landscape of the distribution of mtDNA haplogroups in Europe is a term should not be used in its absolute meaning. It has already been shown (Richards *et al.,* 2002; Richards, 2003) that a deeper phylogenetic analysis allows the revealing of significant differences in the spread of mtDNA sub-haplogroups in Europe – and not only among genetic outliers. We can provide such an example also from our study of hg H genomes in the Latvian sample. Hg H, which is the most frequent hg in all European populations except the Saami, accounted for almost half (43%) of the mtDNA variants in Latvians. Forty-five per cent of the classified hg H genomes belonged to sub-hgs H1 and H5 (see table 3, figure 3, ref. **I**). Sub-hg H1b, which occurs more frequently in Eastern and North-Central Europe (ca. 7% and 5% from the total of hg H, respectively, (Loogväli *et al.,* 2004), was the most abundant type of sub-hg H1 among Latvians (25% of H1, 8% of H), also being frequent among Estonians (see table 3, ref. **I**). The HVS-I sequence motifs characteristic to this sub-hg can often be observed also in the Lithuanian population (9.5% from hg H, Kasperavičiūtė *et al.,* 2004). In the Latvian gene pool, H1b was significantly more frequent than among Eastern Slavs ($p \leq 0.025$). The frequency of H1a in the Latvian population was similar to that of other North-Eastern and Eastern European populations (see table 3, ref. **I**). Significant differences of H1a frequencies were observed between Latvians and Finns ($p \leq 0.025$). Among the latter, 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 ($p \leq 0.001$). 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 (6%; $p \leq 0.025$).

A quarter of mtDNA variants belonged to hg U in the Latvian gene pool. Five sub-hgs – U2, U3, U4, U5 and K – were observed among hg U haplotypes. The sub-hgs U4 and U5 were most abundant, covering about 40% of the hg U gene pool. Interestingly, the frequency of sub-hg U4 in Latvians (8.3%) is among the highest in Europe. Its highest frequencies can be actually observed in Ob-Ugric Khants and Mansis, who live nowadays on the eastern side of the Ural Mountain (Derbeneva *et al.,* 2002b). A relatively high frequency and heterogeneity of hg U4 lineages in the Latvian mtDNA pool may point on early migration and genetic contribution of the first settlements from the Periglacial (northern Ukrainian) refugium around the peak of the LGM (Dolukhanov, 2000, Tambets *et al.,* 2003). The analysis of the mtDNA hg profiles in different European populations (see table 2, ref. **I**) showed that the frequency of hg U4 is significantly higher in Eastern than in Western European populations ($p \leq 0.001$, see details about population groups in the *Methods* section in ref. **I**). In particular, significant differences of U4 frequencies were observed between Latvians and French, as well as between Latvians and Finns ($p \leq 0.001$).The other hg that has a different frequency pattern in Eastern and Western European populations (see table 2, ref. **I**) 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 ($p \leq 0.01$). Hg W, which encompasses *ca* 3% of mtDNA variants of Latvian gene pool, has also some haplotypes, not reported in other populations. Although the frequency (see table 3, appendix 1) as well as the HVS-I and HVS-II diversity of hg W haplotypes of Latvians (0.83) is comparable to those of neighboring populations (*e. g.* 0.96 for Russians; 0.86 for Poles; 0.88 for Germans; but only 0.38 for Finns), the proportion of unique haplotypes is notably high. Three out of seven haplotypes represented in Latvian mtDNA pool does not have exact matches in published datasets (see the *Subjects* section). Two individuals belonged to haplotype 16223-16291-16292, one to haplotype 16051-16223-16292 and one to the haplotype 16179-16292, with probable back-mutation at position 16223. To illustrate the genetic relationships of studied populations, PC analysis based on the frequencies of mtDNA hgs was performed (see figure 2, ref. **I**). To see whether the close genetic relationships of paternal lineages of the Baltic-speakers and Finno-Ugric-speaking Mari, observed by Laitinen *et al.,* (2002), can also be also seen while comparing the maternal lineages of these populations, the mtDNA data of Mari were included, among others, into the PC analysis. The analysis showed

that all European populations, except Finns, French and Mari, fall into one cluster. Latvians and Lithuanians formed a tight cluster with Estonians, Russians, Poles, Germans and Norwegians.
One hundred and twenty six different haplotypes belonging to 10 major hgs were observed in a sample of 351 mtDNAs from the Latvian population (appendix 2; see table 1, ref. **I**). In order to compare the mtDNA haplotype distribution in Latvians with those observed in neighbouring populations we examined the HVS-I sequence variation, also taking into account the information from coding region mutations, where available, in the context of published sequence data. Additional to those populations listed in table 1 (ref. **I**), the comprehensive mtDNA database (for references, see *Subjects and Methods*, ref. **I**) was used for background information. One of our aims was to see whether the Baltic-speaking populations possess mtDNA variants that could reflect the influence of their distinct linguistic background compared to that of their geographical neighbours. The comparison of mtDNA haplotypes from Baltic-speakers revealed that most of the lineages shared between Latvians and Lithuanians are also present in neighbouring populations, in many cases among both Finno-Ugric- and Slavonic speakers, as well as among Western European Indo-European-speakers (see table 1, ref. **I**). Only one lineage from sub-hg I1 with the HVS-I motif 16129–16172–16223–16311–16319 was not found in our database, but was present both in the Latvian (Semigalian) as well as in the Lithuanian (North Žemaičiai) mtDNA pool. The other haplotype from hg H sub-hg H11 with HVS-I motif 16224–16278–16293–16311, shared mostly by Baltic-speakers, has been rarely found in different Eastern European populations (Loogväli *et al.,* 2004) and appears also in Central Asian populations (Comas *et al.,* 1998). Both of these haplotypes are derived from two frequent founder-haplotypes – 16278–16293– 16311 of sub-hg H11 and 16129–16172–16223–16311 of hg I, respectively, which have a wide geographical spread. While the proportion of shared lineages among different populations was found to be quite similar for all populations studied, there are also examples where some derived haplotypes appear to be associated with a specific region or linguistic group. For example, some haplotypes of sub-hgs U2 and U4 that were detected in the Latvian mtDNA pool have so far been found only among Finno-Ugric-speaking populations (ref. **I**). Analogous examples can also be found if one compares the mtDNA pools of Latvians and Slavonic- as well as Germanic-speaking populations (see table 1, ref. **I**). Results of the analysis of the genetic structure (AMOVA) of investigated populations are presented in table 4 (ref. **I**). Samples were grouped according to their linguistic background and on the basis of the geographical location of populations. Although by far the largest fraction of genetic variation was found within populations, still in all groupings the proportion of genetic variation within groups and between individual populations was clearly higher than that in between group comparisons. The largest effect was observed for the Baltic-Finno-Ugric and the smallest for the Baltic-Slavic comparisons. The proportions of the variation between groups and within groups did not show any consistent difference when linguistic grouping and geographic grouping was compared, demonstrating the complexity for mtDNA diversity patterns in the part of Europe under study.
Presented here results confirm that the gene pool of Latvians is characterized by the same package of Western Eurasian mtDNA hgs that encompass about 95% of mtDNA variation in Europe (Richards *et al.,* 1996, 1998; Torroni *et al.,* 1996), with 70% of lineages belonging to hgs H and U (table 3). Eastern Eurasian hgs form only a minute fraction (0.3%) of the Latvian mtDNA pool, represented by a single hg G2a lineage (see table 1, ref. **I**). In Northern Europe, the two main hg M lineage groups are D5b and Z1 that are present in many populations at low frequencies (Tambets *et al.,* 2004). Interestingly, these hgs were observed neither in our sample of 351 Latvians nor among 225 Lithuanians (Kasperavičiūtė *et al.,* 2004; Tambets *et al.,* 2004). In the latter the derivatives of hg M were not sampled, and the Eastern Eurasian package of mtDNA lineages was represented by a single hg A mtDNA. One D5b individual has been described among 545 Estonians (Tambets *et al.,* 2004). This observation is of some interest because the more Northern Finnic-speaking populations, Finns, Saami and Karelians, possess East Asian maternal lineages at frequencies that are also low, but still higher – 2%-5% – than in populations living in the South-eastern Baltic coast (see table 2, ref. **I**; Sajantila *et al.,* 1995;

Tambets *et al.,* 2004). This suggests that East Asian maternal gene flow had, in Eastern Europe, only a negligible impact on populations living in the South-eastern Baltic Sea region, though it reached populations living close to the sub-Arctic fringe of Europe. The HVS-I haplotype-sharing analysis among Baltic-, Germanic-, Slavonic-, and Finno-Ugric-speaking populations (see table 1, ref. **I**) 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 the linguistic affiliations of the latter. These results most likely reflect a deep common origin for the European mtDNA pool (Richards *et al.,* 1996).

5.2. The Saami: their position as so-called genetic "outliers" among European populations and a case study of haplogroup U5 in Latvian population (Ref. II)

Because of the persistent ambiguity in the origin of the Saami population, their haploid genomes have been investigated in a comprehensive Eurasian context (ref. **II**). As the mtDNA heritage and variation of the Saami is not the topic of the present dissertation, this aspect of the ref. **II** is described herein only briefly and the main emphasis here is on the variation of the sub-haplgroup U5b1b among the Saami in a context of it variation in Eurasian populations. Besides, some peculiarities of haplogroup U5 distribution in different ehtnolinguistic groups of the Latvian population are also discussed herein.

5.2.1. mtDNA analysis of the Saami and phylogeography of haplogroup U5b1b

Analysis of Swedish Saami, combined with the reanalysis of previously published sequences of other Saami populations from Finland and Norway (Sajantila *et al.,* 1995; Dupuy and Olaisen, 1996; Delghandi *et al.,* 1998) showed that the "outlying" status of the Saami is caused by relative haplogroup proportions in their mtDNA pool, not by distinctive phylogeographical affiliation of their maternal lineages (fig. 2A in ref. **II**). As a specific novelty, we have shown that there is little, if any, historic gene flow from Samoyedic- and Ugric-speaking populations from Siberia to Fennoscandia alongside the Arctic zone. The Saami mtDNA pool consists predominantly of two haplogroups that are widely spread in Europe: V and U5 (table 1 and 2 in ref. **II**; fig. 2 in Tambets *et al.,* 2001; table 1 in Torroni *et al.,* 2001b). These haplogroups together cover more than 90% of the Saami mtDNA lineages. The rest of the variation is shared by European-specific hgs H, W and T (fig.1 in ref. **II**). Only a small fraction of mtDNA variants shows the contribution from eastern Eurasia (hgs D5 and Z1, table 1 in ref. **II**). The most plausible explanation for the unconventional frequency pattern of mtDNA haplogroups among the Saami is that genetic drift (bottleneck and founder-effect) has had a major role in shaping the mtDNA pool of this small European populations.

Most of the U5b lineages spread among the Saami possess the so-called "Saami-motif" – a combination of three transitions, 16144 (T to C), 16189 (C to T) and 16270 (C to T) in their HVS-I sequences, which was previously believed to be restricted only to the Saami – occasional findings from neighbouring populations, in the Karelians and Finns, were explained by their admixture with the Saami (Sajantila *et al.,* 1995; Meinilä *et al.,* 2001). The informative coding region nps (table 4 in ref. **II**) in U5b topology have been analyzed and showed that "the Saami variant" of U5b (U5b1b1), predominant among the Saami, is widely spread among different eastern European populations, extending, at very low frequencies, also to western Europe and to the Caucasus (fig. 3B and table 1 in ref. **II**). On the other hand, U5b1b1 forms a sub-branch of a lineage cluster, defined here as U5b1b (5656G, 7385G, 10927C). Phylogeographic analysis of U5b1b revealed that this haplogroup, similarly to V, another major haplogroup in Saami (see also Torroni *et al.,* 2001b), has a greater diversity in western than in Eastern Europe, supporting a scenario, according to which they both started their expansion after the LGM from western parts of Europe, possibly from Franco-Cantabrian refugium. Meanwhile, the analysis of different European populations suggest that these haplogroups have likely reached northern Fennoscandia

not along the Atlantic coast of the Scandinavian Peninsula, but rather via the "eastern route" (fig.4 in ref. **II**), as the haplotypes of hgs U5 and V of the Saami are well present in eastern European populations. The phylogeographic pattern of U5b1b1 suggests that this particular sub-clade might have arisen in East Europe.

5.2.2. Phylogenetic study of haplogroup U5 in Latvian population

Sub-haplogroup U5 forms 11.7% from the whole maternal lineages of the Latvians and is depicted by two main sub-haplogroups, U5a and U5b (table 3). However, a minute fraction of observed mtDNA variation compiles U5b sequences (31.7% from all U5 samples or 3.7% from all mtDNA genomes). Sub-hg U5b is represented by three sub-branches (U5b*, U5b1 and U5b1b), from them U5b1b was found only in two individuals from South-Eastern and Eastern regions of Latvia (see table 4, ref. **II**). Despite being found among the Saami, we have not observed the presence of Ub1b1 lineages in studied Latvians, meanwhile, among another Baltic-speaking population, Lithuanians, this particular mtDNA variant was detected in three individuals bearing two HVS-I haplotypes 16144-16189-16270 and one-step derivate from it 16104-16144-16189-16270 (Kasperavičiūtė *et al.,* 2004). Here, we also provide additional evidence of presence of the rare haplotype 16192-16311 (sub-hg U5b*), which is probably generated by the back-mutation at nps 16270 in HVS-I. This mtDNA lineage was found at low frequencies in different populations from South Europe, Scandinavia and Eastern Europe (see table 4, ref. **II**), and was the second frequent mtDNA haplotype (23.1%) from U5b genomes in Latvians. Interestingly, performing the micro-phylogeographic analysis of U5b lineages at the regional level of Latvia, we have noticed that this sub-branch of haplogroup U5 was more abundant in South-West, South-East and Eastern parts of Latvia, but was absent in one region – Northern Curonia (table 5).

Table 5. mtDNA sub-haplogroup U5 HVS-I haplotypes of Latvian sub-populations

HVS-I haplotype (-16 000)	sub-hg	SE	LE	NC	SC
114ca 192 256 270 286cg 292 294	**U5a**		1		
114ca 192 256 270 292 294	**U5a**	1	2		
114ca 192 256 270 294	**U5a**				1
114ca 192 256 270 294 311	**U5a**	1	1		
134 192 256 270 291	**U5a**		1		
174 256 270	**U5a**				1
189 256 270	**U5a**				1
192 248 256 270 291 294 311	**U5a**				1
192 256 270 292 294	**U5a**		1		
192 256 270 311	**U5a**		1		
192 256 270 320	**U5a**	2		1	
256 270	**U5a**	1	3	2	
256 270 294	**U5a**	1	1		
093 192 256 270 291	**U5a**		2	1	1
192 311	**U5b**		2		1
189 270	**U5b1**	1	2		1
189 248 270	**U5b1**		1		
192 270	**U5b1**	1			
189 192 270	**U5b1**		2		
093 189 270	**U5b1b**		1		
093 129 189 270	**U5b1b**		1		

Population codes are given in alphabetical order as follows: LE – Lettigalians; NC – Northern Curonians; SC – Southern Curonians; SE – Semigalians.

Sub-haplogroup U5a is notably more frequent (68.3% from all U5 samples or 7.9% from all mtDNA genomes) and diverse among Latvians (0.93 *versus* 0.88 for sub-hg U5b). Within sub-hg U5a one part of HVS-I haplotypes detected in Latvians possess a transversion at np 16114 (C to A), and a core haplotype is formed by four additional substitutions at nps 16192, 16256, 16270 and 16294. This founder mtDNA lineage is widely spread across different populations from the Near East (Richards *et al.,* 1998; 2000; Torroni *et al.,* 1998) and Caucasus area (Metspalu *et al.,* 1999), Western and Eastern Europe as well as from Volga-Ural region (*e.g.* Pereira *et al.,* 2000; Helgason *et al.,* 2001; Malyarchuk *et al.,* 2001; 2002; Bermisheva *et al.,* 2002). However, the main stress was on the results associated with local distribution of previously mentioned U5a lineage. We found out non-uniform distribution of it in different regions of Latvia, again, as for U5b sub-clades, sub-hg U5a lineages with transverion at np 16114CA were obviously restricted to the South-Eastern regions of Latvia and completely absent in Northern Curonia. In order to understand its spread pattern in western Eurasia a broad phylogeographic analysis and more extensive knowledge about coding region variation of all U5a sub-clades is still needed, because many of them are thus defined only by fast evolving nps in HVS-I.

The absence of sub-hg U5b and low diversity of sub-branches of U5a in Northern Curonia in comparison with other regions of Latvia may be due to several reasons; one may explain, U5b and U5a lineages that probably were reached Latvia via "eastern route" in some other manner had not been spread in North-western part of Latvia, in such a way pointed on rather different ethnolinguistic component of this region. If the territory of the Baltic peoples was threatened from the East and South by eastern and western Slavs, then on the West, the Baltic tribes (Prussians and Curonians) where the object of attempted conquest by the Scandinavians. With respect to ethnic processes in what is now Latvia, we must note that Baltic tribes regrouped, seeking to expand their lands to the North, where Baltic Finns (Livs and Estonians) lived. In the middle of the first millennium AD, Semigallians and Selonians stopped burying their dead in burial mounds, instead choosing level skeletal graves where the dead were buried unburned. This process developed differently in the western part of Latvia territory, where Curonians began to use level skeletal graves in the first centuries AD, replacing them with fire graves in the 9th and 10th century (Mugurevics, 1997).

Overall, U5 is a "prototype" western Eurasian lineage cluster with a coalescence age around 45,000-55,000 YBP, and its phylogenetic tree does not suggest a star-like expansion from the founder (*e.g.* Richards *et al.,* 2000; Finnilä *et al.,* 2001). However, Tambets *et al.,* (2003) have revealed the presence of about a dozen putative sub-founders, most of which exhibit nice star-like expansions. More importantly, it has been found out that almost all of them exhibit coalescence ages around 11,000-13,000 YBP and only a few, like "the Saami U5", seem to have started to expand significantly more recently. This nearly synchronous series of coalescence ages makes sense: it is much easier to imagine (specifically for a such an ancient branch) that an expansion phase hit all U5 twigs and limbs nearly simultaneously, than to assume a complicated pattern of a dozen or so widely irregular beginnings (Tambets *et al.,* 2003).

5.3. Y-chromosomal variation of the Latvians

Although the Latvian Y-chromosomal variation is not the topic of this PhD dissertation, it seems appropriate to discuss it here briefly. Studies of Y-chromosome markers have revealed that approximately one third to a half of all Y-chromosomes found in Latvians and Lithuanians belong to hg N3, defined by the mutation TatC (Zerjal *et al.,* 1997, 2001; Lahermo *et al.,* 1999; Laitinen *et al.,* 2002; Tambets *et al.,* 2004). This high frequency of hg N3, combined with its considerable diversity of microsatellite haplotypes, has also been found among Estonians (30–35%) and Volga-Finnic-speaking populations (20– 50%) (Rosser *et al.,* 2000; Zerjal *et al.,* 2001; Tambets *et al.,* 2004). The proportion of this variant of Y chromosomes drops drastically in the geographical neighbours of Latvians and Lithuanians – among Poles (2%), and is much less frequent also among Slovaks (3%), Ukrainians (6%) and Russians (8–14%) (Rosser *et al.,* 2000;

Tambets *et al.,* 2004). The sharp decline of the frequency of hg N3 from 48% among the Finno-Ugric-speaking Saami to less than 8% among the Norwegians and Swedes can also be observed in Northern Scandinavia (Rosser *et al.,* 2000; Zerjal *et al.,* 2001; Tambets *et al.,* 2004). Thus, based on N3 frequency distribution, Latvians and Lithuanians are closer to the adjacent Finno-Ugric-speakers than to their Slavonic-speaking neighbours, with whom they share Indo-European linguistic proximity. This finding has been interpreted as evidence of a common origin for Baltic- and Finnic-speakers (Laitinen *et al.,* 2002). Recently, Kasperavičiūtė *et al.,* (2004) found, similarly to Zerjal *et al.,* (2001), that the microsatellite haplotype patterns within hg N3 are different among Lithuanians and Estonians. They noted, however, that it is unclear whether the observed differences suggest different source populations, as proposed by Zerjal *et al.,* (2001), or rather more recent random genetic drift. Meanwhile, it is interesting to note that the calculations of Kasperavičiūtė *et al.,* (2004) suggested very similar expansion times, around 7000 – 8000 YBP, for the Lithuanian and Estonian hg N3 Y chromosomes, which are probably also applicable to Latvians. Therefore, it is indeed possible that the spread of hg N3 among the ancestral populations of Estonians, on the one hand, and the Baltic-speaking populations on the other, predate the advent of the Neolithic age in the East Baltic and may be part of the post-LGM re-colonisation of the region. An ancient language shift from Uralic to Indo-European among the Baltic-speakers has been suggested, linked to an earlier arrival of agriculture to the ancestors of the present-day Latvians and Lithuanians (Wiik, 2000). The genetic proximity of the Baltic-speaking populations with the Volga-Finnic Mari, proposed based on Y-chromosomal diversity of those populations by Laitinen *et al.,* (2002), most likely reflects simply an ancient, largely common heritage of Eastern European populations, rather than a specific link between the two populations – our analysis of the mtDNA hg frequency profiles (see fig. 2, ref. **I**) does not cluster Mari and Latvians closely together, and the comparison of mtDNA HVS-I haplotypes in these populations shows that they predominantly share only founder lineages, while differing for more derived haplotypes (see table 1, ref. **I**).

It is also interesting to note that the populations in the East Baltic are the northern boundary of the spread of the "Adriatic" NRY hg I1b*-P37, present at very low frequencies both in Latvians and Estonians, while considerably more frequent in Byelorussians, Ukrainians, southern Russians and, in particular, in the Balkans (Rootsi *et al.,* 2004). On the other hand, though there is no geographic border between Estonia and Latvia, there is a three-fold southwards frequency drop (14.8% *versus* 4.7%) of hg I1a*-M253, which is particularly frequent all over Scandinavia, including among the Saami (Rootsi *et al.,* 2004). Its moderate presence in Estonia is probably due to the long-term presence of Swedish settlements in Estonian islands and in the north-western coastal areas. The observed maternal gene flow (migration of females) contrasts with an apparent lack of westward flow of hg N3 Y chromosomes (migration of males). The Baltic-speaking populations largely share a mosaic of frequencies that, in Europe, brings them together specifically with the Finno-Ugric-speaking populations, both those living in the Baltic area and those living in the Volga Basin.

6. CONCLUSIONS

The results of the present study can be summarized as follows:

1. The experimental basis of the current thesis relies on the analysis of the mtDNA variation in the Latvian population as a whole and it subpopulations (sample size – 351 individuals). For comparison, the mtDNA database of 11,236 individuals, consisting of published mtDNA data as well as unpublished mtDNA data from 4,732 individuals from different Eurasian populations were used as background information for the analysis. A DNA fragment, encompassing the mtDNA HVS-I was amplified and sequenced in all samples; HVS-II sequences were determined only for selected samples from hgs U4, H and W. To confirm the haplogroup affiliations of mtDNA sequences, hierarchical RFLP analysis was performed using 17 restriction endonucleases.

2. We have shown that mtDNA haplogroup frequencies distribution and their heterogeneity across Latvian sub-populations were non-uniform. Statistically significant differences in haplogroup frequencies between the eastern and the western/central parts were seen within four sub-lineages of one of the most diverse and abundant haplogroup among Latvians, U (U2, U4, U5 and K, correspondingly).

3. The region-specific differences in the Latvian mtDNA pool were mostly observed between Western (Northern Curonia) and Eastern (Lettigalia) part of Latvia and were affirmed further by several statistical approaches.

4. The genetic trace of the forced internal migration movement from eastern/south-eastern region (Lettigalia) towards the north-western and central parts of the country can be seen based on mtDNA variation, and is in accordance with archaeological data. The sensitivity of mtDNA to the population substructure makes it a good marker for detecting regional variation.

5. We have shown that the gene pool of Latvians is characterized by the same package of Western Eurasian mtDNA haplogroups that encompass about 95% of mtDNA variation in Europe, with 70% of lineages belonging to haplogroups H and U. Eastern Eurasian haplogroups form only a minute fraction (0.3%) of the Latvian mtDNA pool, represented by a single hg G2a lineage.

6. The uniformity of the genetic landscape of the distribution of mtDNA haplogroups in Europe is a term should not be used in its absolute meaning. Recently, it has already been shown that a deeper phylogenetic analysis allows the revealing of significant differences in the spread of mtDNA sub-haplogroups in Europe – and not only among genetic outliers. We can provide such an example also from our study of hg H genomes in the Latvian sample.

7. A relatively high frequency and heterogeneity of hg U4 lineages in the Latvian mtDNA pool may point on early migration and genetic contribution of the first settlements from the Periglacial (northern Ukrainian) refugium around the peak of the LGM.

8. The HVS-I haplotype-sharing analysis among Baltic-, Germanic-, Slavonic-, and Finno-Ugric-speaking populations have 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 the linguistic affiliations of the latter.

9. Our results confirm that all sub-populations of the Saami are characterized by a restricted variation of maternal lineages predominantly European in origin (haplogroups V and U5). Less than 5% of mtDNA variants, spread among the Saami, belong to eastern Eurasian-specific haplogroups.

10. The observed maternal gene flow (migration of females) contrasts with an apparent lack of westward flow of hg N3 Y chromosomes (migration of males). The presence of Y-chromosomal lineages (R1b and I1b hgs) at negligible frequencies among Latvians points on minor genetic components or insignificant gene flow from Western Europe after the Last Glacial Maximum (LGM) (probably, from Franco-Cantabrian refugia). Based on Y-chromosomal variation the Baltic-speaking populations (Latvians and Lithuanians) largely share a mosaic of frequencies that, in Europe, brings them together specifically with the Finno-Ugric-speaking populations, both those living in the Baltic area and those living in the Volga Basin.

REFERENCES

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogvali EL, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. Am J Hum Genet 75: 910-918

Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferrreira AC, Bandelt H-J, Pens SD, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67: 444-461

Ammerman AJ, Cavalli-Sforza LL (1984) The Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton, pp. 200

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290: 457-465

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23: 147

Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics 103: 287-312

Awdalla P, Eyre-Walker A, Smith JM (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science 286: 2524-2525

Baasner A, Schafer C, Junge A, Madea B (1998) Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. Forensic Science International 98: 169-178

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, Wallace DC (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics 130: 139-152

Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, Papiha SS, Villems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, Jorde LB (2001) Genetic evidence on the origins of Indian caste populations. Genome Research 11: 994-1004

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141: 743-753

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

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

Barbujani G, Sokal RR (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. Proc Natl Acad Sci USA 87: 1816-1819

Barker G (1985) Prehistoric Farming in Europe. Cambridge University Press, Cambridge, pp. 493

Batini C, Coia V, Battaggia C, Rocha J, Pilkington MM, Spedini C, Comas D, Destro-Bisol G, Calafell F (2006) Phylogeography of the human mitochondrial L1c haplogroup: Genetic signatures of the prehistory of Central Africa. Mol Phylogenet Evol (in press)

Behar DM, Hammer MF, Garrigan D, Villems R, Bonne-Tamir B, Richards M, Gurwitz D, Rosengarten D, Kaplan M, Della Pergola S, Quintana-Murci L, Skorecki K (2004) MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. Eur J Hum Genet 12; 255-364

Beckman L, Kallstenius P (1959) Blood groups of Estonians, A1A2B0 and Rh. J Hered 6: 287–289

Beckman L, Sikstrom C, Mikelsaar AV, Krumina A, Ambrasiene D, Kučinskas V, Beckman G (1998) Transferrin variants as markers of migrations and admixture between populations in the Baltic Sea region. Hum Hered 48: 185-191

Beckman L, Sikstrom C, Mikelsaar A, Krumina A, Kučinskas V, Beckman G (1999) alpha1 antitrypsin (PI) alleles as markers of Westeuropean influence in the Baltic Sea region. Hum Hered 49: 52-55

Beleza S, Gusmao L, Amorim A, Carracedo A, Salas A (2005) The genetic legacy of western Bantu migrations. Hum Genet 117: 366-375

Bensasson D, Zhang D, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol 16: 314-321

Bermisheva M, Tambets K, Villems R, Khusnutdinova E (2002) Diversity of mitochondrial DNA haplotypes in ethnic populations of the Volga-Ural region of Russia. Mol Biol (Mosk) 36: 990-1001

Blanc H, Chen KH, D'Amore MA, Wallace DC (1983) Amino acid change associated with the major polymorphic Hinc II site of Oriental and Caucasian mitochondrial DNAs. Am J Hum Genet 35: 167-176
Blanco I, de Serres FJ, Fernandez-Bustillo E, Lara B, Miravilles M (2006) Estimated numbers and prevalence of PI*S and PI*Z alleles of alpha1-antitrypsin deficiency in European countries. Eur Respir J 27: 77-84
Bogenhagen DF (1999) Repair of mtDNA in vertebrates. Am J Hum Genet 64: 1276-1281
Brega A, Gardella R, Semino O, Morpurgo G, Astaldi Ricotti GB, Wallace DC, Santachiara- Benerecetti AS (1986) Genetic studies on the Tharu population of Nepal: restriction endonuclease polymorphisms of mitochondrial DNA. Am J Hum Genet 39: 502-512
Brown MD, Hosseini SH, Torroni A, Bandelt H-J, Allen JC, Schurr TG, Scozzari R, Cruciani F, Wallace DC (1998) mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? Am J Hum Genet 63: 1852-1861
Brown WM, George M, Jr., Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76: 1967-1971
Brown WM (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc Natl Acad Sci U S A 77: 3605-3609
Cali F, Le Roux MG, D'Anna R, Flugy A, De Leo G, Chiavetta V, Ayala GF, Romano V (2001) MtDNA control region and RFLP data for Sicily and France. Int J Legal Med 114: 229-231
Calloway CD, Reynolds RL, Herrin GL Jr, Anderson WW (2000) The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. Am J Hum Genet 66: 1384-1397
Cann RL, Brown WM, Wilson AC (1982) Evolution of human mitochondrial DNA: a preliminary report. Prog Clin Biol Res 103: 157-165
Cann RL, Brown WM, Wilson AC (1984) Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics 106: 479-499
Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325: 31-36
Cann RL (2001) Genetic clues to dispersal in human populations: retracing the past from the present. Science 291: 1742-1748
Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, pp. 1088
Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am J Hum Genet 57: 133-149
Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC (2000) mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. Am J Hum Genet 66: 1362-1383
Chikhi L, Destro-Bisol G, Bertolle G, Pascali V, Barbujani G (1998) Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. Proc Natl Acad Sci USA 95: 9053-9058
Chikhi L, Nichols RA, Barbujani G, Beaumont MA (2002) Y genetic data support the Neolithic demic diffusion model. Proc Natl Acad Sci USA 99: 11008-11013
Clark JD, Beyene Y, WoldeGabriel G, Hart WK, Renne PR, Gilbert H, Defleur H Suwa G, Katoh S, Ludwig KR, Boisserie JR, Asfaw B, White TD (2003) Stratigraphic, chronological and behavioural contexts of Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature 423: 747–752.
Coia V, DestropBisol G, Verginelli F, Battaggia C, Boschi I, Cruciani F, Spedini G, Comas D, Calafell F (2005) Brief communication: mtDNA variation in North Cameroon: lack of Asian lineages and implication for back migration from Asia to sub-Saharan Africa. Am J Phys Anthropol 128: 678-681
Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Martinez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D, Pettener D, Bertranpetit J (1998) Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. Am J Hum Genet 63: 1824-1838
Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, Bertranpetit J (2004) Admixture, migration, and dispersals in Central Asia: evidence from maternal DNA lineages. Eur J Hum Genet 11: 1-10
Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum Genet 60: 331-350
Cox DW (1995) AAT deficiency. The metabolic and molecular bases of inherited disease. New York: McGrawhil, 4125-4158
Crespillo M, Luque JA, Paredes M, Fernandez R, Ramirez E, Valverde JL (2000) Mitochondrial DNA sequences for 118 individuals from northeastern Spain. Int J Legal Med 114: 130-132
Darlu P, Tassy P (1987) Disputed African origin of human populations. Nature 329: 111-112

De la Chapelle A, Wright FA (1998) Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. Proc Natl Acad Sci USA 95: 12416-12423
Delghandi M, Utsi E, Krauss S (1998) Saami mitochondrial DNA reveals deep maternal lineage clusters. Hum Hered 48: 108-114
Denaro M, Blanc H, Johnson MJ, Chen KH, Wilmsen E, Cavalli-Sforza LL, Wallace DC (1981) Ethnic variation in Hpa 1 endonuclease cleavage patterns of human mitochondrial DNA. Proc Natl Acad Sci USA 78: 5768-5772
Denisova R (1975) Antropologiya drevnih baltov. Riga, Zinātne,153 lpp.
Denisova R (1994) Zvejnieku akmens laikmeta kapulauka veidošanās vēsture saistībā ar iegūtām radioaktīvā oglekļa datējumiem (History of the formation of the Zvejnieki Stone Age burial field according to obtained 14C datings). Latvijas Vēstures Institūta Žurnāls, No. 4: 37-46
Denisova R (1996) Zvejnieku akmens laikmeta kapulauks. Iekšējā struktūra un hronoloģija (The Zvejnieki Stone Age burial field: Internal structure and Chronology). Latvijas Vēstures Institūta Žurnāls, No. 4: 47-54
Denisova R (1997) The most ancient populations of Latvia. Journal of the Institute of Latvian History 3: 5-15
Derbeneva OA, Starikovskaia EB, Volod'ko NV, Wallace DC, Sukernik RI (2002a) [Mitochondrial DNA variation in Kets and Nganasans and the early peoples of Northern Eurasia]. Genetika 38: 1554-1560
Derbeneva OA, Starikovskaya EB, Wallace DC, Sukernik RI (2002b) Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. Am J Hum Genet 70: 1009-1114
Derenko MV, Dambueva IK, Malyarchuk BA, Dorzhu CM, Zakharov IA (1999) Structure and diversity of the mitochondrial gene pool in indigenous populations of Tuva and Buryatia based on restriction polymorphism data. Genetika 35: 1706-1712
Derenko MV, Grzybowski T, Malyarchuk BA, Czarny J, Miscicka-Sliwka D, Zakharov IA (2001) The presence of mitochondrial haplogroup X in Altaians from South Siberia. Am J Hum Genet 69: 2371-241
Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Miscicka-Sliwka D, Wozniak M, Zakharov IA (2003) Diversity of mitochondrial DNA lineages in South Siberia. Ann Hum Genet 67: 391-411
Di Rienzo A,Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88: 1597-1601
Dimo-Simonin N, Grange F, Taroni F, Brandt-Casadevall C, Mangin P (2000) Forensic evaluation of mtDNA in a population from south west Switzerland. Int J Legal Med 113: 89-97
Dolukhanov PM, Gey NA, Miklyayev AM, Mazurkiewicz AN (1989) Rudnya-Serteya, a stratified site in the Upper Duna basin. Fennoscandia Archaeologica, VI
Dolukhanov PM (2000) "Prehistoric revolutions" and languages in Europe. In: Künnap A (ed) The roots of peoples and languages of Northern Eurasia: II and III. University of Tartu. Division of Uralic Languages, Societas Historiae Fenno-Ugricae, Tartu, pp 71-84
Dubut V, Chollet L, Murail P, Cartault F, Beraud-Colomb E, Serre M, Mogentale-Profizi N (2004) mtDNA polymorphisms in five French groups: importance of regional sampling. Eur J Hum Genet 12: 293-300
Dupuy BM, Olaisen B (1996) MtDNA sequences in the Norwegian Saami and main population. In: Carracedo A, Brinkmann B, Bär W (eds) Advances in forensic haemogenetics. 6. Springer-Verlag, Berlin, Heidelberg, New York, pp 23-25
Eisensmith RC, Goltsov AA, O'Neill C, Tyfield LA, Schwartz EI, Kuzmin AI, Baranovskaya SS, Tsukerman GL, Treacy E, Scriver CR (1995) Recurrence of the R408W mutation in the phenylalanine hydroxylase locus in Europeans. Am J Hum Genet 56: 278-286
Elson JL, Samuels DC, Turnbull DM, Chinnery PF (2001) Random intracellular dirft explains the clonal expansion of mitochondrial DNA mutations with age. Am J Hum Genet 68: 802-806
Elson JL, Turnbull DM, Howell N (2004) Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. Am J Hum Genet 74: 229-238
Excoffier L, Langaney A (1989) Origin and differentiation of human mitochondrial DNA. Am J Hum Genet 44: 73-85
Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distancesamong DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479-491
Excoffier L, Yang Z (1999) Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Mol Biol Evol 16: 1357-1368
Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50
Eyre-Walker A, Smith NH, Smith JM (1999) How clonal are human mitochondria? Proc E Soc Lonf B Biol Sci 266: 477-483

Fedorova SA, Bermisheva MA, Villems R, Maksimova NR, Khusnutdinova EK (2003) Analysis of mitochondrial DNA lineages in Yakuts. Mol Biol (Mosk) 37: 643-653
Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. Journal of Mol Evol 17: 368-376
Finnilä S, Lehtonen MS, Majamaa K (2001) Phylogenetic network for European mtDNA. Am J Hum Genet 68: 1475-1484
Finnilä S, Majamaa K (2001) Phylogenetic analysis of mtDNA haplogroup TJ in a Finnish population. J Hum Genet 46: 64-69
Fitch W (1977) On the problem of discovering the most parsimonious tree. Am Nat 111: 1169-1175
Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59: 935-945
Forster P, Torroni A, Renfrew C, Rohl A (2001) Phylogenetic Star Contraction Applied to Asian and Papuan mtDNA Evolution. Mol Biol Evol 18: 1864-1881
Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D (2004) Assessing the impact of population stratification on genetic association studies. Nat Genet 36: 388-393
Fucharoen G, Fucharoen S, Horai S (2001) Mitochondrial DNA polymorphisms in Thailand. J Hum Genet 46: 115-125
Futuyma, Douglas J (2005) On Darwin's Shoulders. Natural History 114: 64–68.
Gerber AS, Loggins R, Kumar S, Dowling TE (2001) Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes? Ann Rev Genet 35: 539-566
Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci USA 77: 6715-6719
Gimbutas M (1963) The Balts. Thames and Hudson, London, Ancient peoples and places, pp. 288
Gimbutas M (1970) Prot-Indo-European Culture: The Kurgan Culture during the fifth, fourth, and third millennia B.C. In: Cardona G, Hoenigswald HM, Senn A (eds) Indo-European and Indo-Europeans. University of Pennsylvania Press, Philadelphia, pp 155-195
Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA (2006) Whole mtDNA Genome Sequence Analysis of Ancient African Lineages. Mol Biol Evol (in press)
Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP (1998) Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9: 585-598
Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 12: 334-345
Gray RD, Atkinson QD (2003) Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature 426: 435-439
Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T, Zegura SL (1997) The geogrpahic distribution of human Y chromosome variation. Genetics 145: 787-805
Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15: 427-441
Harpending HC, Sherry ST, Rogers AR, Stoneking M (1993) The genetic structure of ancient human populations. Curr Anthropol 34: 483-496
Hasegawa M, Di Rienzo, Kocher TD, Wilson AC (1993) Toward a more accurate time scale for the human mitochondrial DNA tree. J Mol Evol 37: 347-354
Hasegawa M, Cao Y, Yang Z (1998) Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. Mol Biol Evol 15: 1499-1505
Hayashida K, Omagari K, Masuda J, Hazama H, Kadokawa Y, Ohba K, Kohno S (2005) The sperm mitochondria-specific translocator has a key role in maternal mitochondrial inheritance. Cell Biol Int 29: 472-481
Hedges SB, Kumar S, Tamura K, Stoneking M (1992) Human origins and analysis of mitochondrial DNA sequences. Science 255: 737-739

Helgason A, Sigurdadottir S, Gulcher J, Ward R, Stefanson K (2000) mtDNA and the origins of the Icelanders: deciphering signals of recent population history. Am J Hum Genet 66: 999-1016

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

Helgason A, Yngyadottir B, Hrafnkelsson B, Gulcher J, Stefansson K (2005) An Icelandic example of the impact of population structure on association studies. Nat Genet 37: 90-95

Henze K, Martin W (2003) Evolutionary biology: essence of mitochondria. Nature 426: 127-128

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet 70: 1152-1171

Hines J (1991) Reviews. Nor Archael Rev 24: 49-54

Hofmann S, Jaksch M, Bezold R, Mertens S, Aholt S, Paprotta A, Gerbitz KD (1997) Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. Hum Mol Genet 6: 1835-1846

Horai S, Gojobori T, Matsunaga E (1984) Mitochondrial DNA polymorphism in Japanese. I. Analysis with restriction enzymes of six base pair recognition. Hum Genet 68: 324-332

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92: 532-536

Housley RA, Gamble CS, Street M, Pettitt P (1997) Radiocarbon evidence for the late glacial human recolonisation of northern Europe. Proc Prehist Soc 63: 25-54

Ilonen J, Koskinen S, Nejentsev S, Sjoroos M, Knip M, Schwartz EI, Adojaan B, Kovalchuk L, Sochnevs A (1997) HLA-DQB1*0304-DRB1*0408 haplotype associated with insulin-dependent diabetes mellitus in populations in the eastern Baltic region. Tissue Antigens 49: 532-534

Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408: 708-713

Ingman M, Gyllensten U (2001) Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. J Hered 92: 454-461

Innan H, Nordborg M (2002) Recombination or mutational hot spots in human mtDNA? Mol Biol Evol 19: 1122-1127

Jaanits K (1990) The Origin and Development of the Kunda Culture: Finno -Ugric Studies in Archaeology, Anthropology and Ethnography. Paper presented at the 6th International Finno-Ugric Congress. Tallinn: Institute of History.

Jin L, Underhill PA, Doctor V, Davis RW, Shen P, Cavalli-Sforza LL, Oefner PJ (1999) Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. Proc Natl Acad Sci USA 96: 3796-3800

Jobling MA, Hurles ME and Tyler-Smith C (2004) Human Evolutionary Genetics: origins, peoples and disease. London/New York: Garland Science Publishing, pp. 523

Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL (1983) Raidation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. J Mol Evol 19: 255-271

Jorde LB, Bamshad M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288: 1931

Kasnauskiene J, Giannattasio S, Lattanzio P, Cimbalistiene L, Kučinskas V (2003) The molecular basis of phenylketonuria in Lithuania. Hum Mutat 21: 398

Kasperavičiūtė D, Kučinskas V, Stoneking M (2004) Y chromosome and mitochondrial DNA variation in Lithuanians. Ann Hum Genet 68: 438-452

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217: 264-626

Kittles RA, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, Goldman D, Long JC (1999) Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. Am J Phys Anthropol 108: 381-399

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R (1999a) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr Biol 9: 1331-1334

Kivisild T, Kaldma K, Metspalu M, Parik J, Papiha SS, Villems R (1999b) The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the Old World. In: Papiha SS, Deka R, Chakraborty R (eds) Genomic Diversity. Kluwer Academic/Plenum Publishers, pp 135-152

Kivisild T, Papiha SS, Rootsi S, Parik J, Kaldma K, Reidla M, Laos S, Metspalu M, Pielberg G, Adojaan M, Metspalu E, Mastana SS, Wang Y, Golge M, Demirtas H, Schnakenberg E, de Stefano GF, Geberhiwot T, Claustres M, Villems R (2000) An Indian Ancestry: a key for understanding human diversity in Europe and beyond. In: Renfrew C, Boyle K (eds) Archaeogenetics: DNA and the populations prehistory of Europe. McDonald Institute for Archaeological Research University of Cambridge, Cambridge, pp 267-279

Kivisild T, Villems R (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288: 1931

Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt H-J, Villems R (2002) The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol 19: 1737-1751

Kivisild T, Rootsi S, Metspalu M, Metspalu E, Parik J, Kaldma K, Usanga E, Mastana S, Papiha SS, Villems R (2003a) The Genetics of Languages and Farmign Spread in India. In: Bellwood O, Renfrew C (eds) Examining the farming/language dispersal hypothesis. The McDonald Institute for Archaerological Research University of Cambridge, Cambridge, pp 215-222

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Golge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R (2003b) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet 72: 313-332

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R (2004) Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet 75: 753-770

Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman M, Cavalli-Sforza LL, Oefner PJ (2006) The role of selection in the evolution of human mitochondrial genomes. Genetics 172: 373-387

Kolman CJ, Bermingham E (1997) Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama. Genetics 147: 1289-12302

Kong QP, Yao YG, Liu M, Shen SP, Chen C, Zhu CL, Palanichamy MG, Zhang YP (2003a) Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. Hum Genet 21: 391-405

Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, Zhang YP (2003b) Phylogeny of East Asian Mitochondrial DNA Lineages Inferred from Complete Sequences. Am J Hum Genet 73: 671-676

Kopitin VF (1979) Mezolit poseleniya Gorki v Posožye. Kratkiye soobčeniya AN SSSR, vip. 157

Krings M, Stone A, Schmitz RW, Krainitzki J, Stoneking M, Pääbo S (1997) Neanderthal DNA sequences and the origin of modern humans. Cell 90: 19-30

Kruger J, Vogel F (1989) The problem of our common mitochondrial mother. Hum Genet 82: 308-312

Krumina A, Kroshkina V, Krumina L (2001) Cystic fibrosis mutation dF508 in the Latvian population. Proceedings of Riga Stradins University: 161-166

Kumar S, Hedrick P, Dowling T, Stoneking M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288: 1931

Lahermo P, Sajantila A, Sistonen P, Lukka M, Aula P, Peltonen L, Savontaus ML (1996) The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. Am J Hum Genet 58: 1309-1322

Lahermo P, Savontaus ML, Sistonen P, Beres J, de Knijff P, Aula P, Sajantila A (1999) Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. Eur J Hum Genet 7: 447-458

Lahr MM, Foley RA (1998) Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. Am J Phys Anthropol Suppl: 137-176

Laitinen V, Lahermo P, Sistonen P, Savontaus ML (2002) Y-chromosomal diversity suggests that Baltic males share common Finno-Ugric-speaking forefathers. Hum Hered 53: 68-78

Landsteiner K (1901) Ueber Agglutinationserscheinungen normalen menschlichen Blutes. Wien. Klin. Wochenschr. 14: 1132–1134 [Translation: On agglutination phenomena of normal human blood, in S. H. BOYER (Editor), 1963, *Papers on Human Genetics*, pp. 27–31. Prentice-Hall, Englewood Cliffs, NJ.]

Lappalainen T, Koivumaki S, Salmela E, Huoponen K, Sistonen P, Savontaus ML, Lahermo P (2006) Regional differences among the Finns: a Y-chromosomal perspective. Gene 376: 207-215

Larruga JM, Diez F, Pinto FM, Flores C, Gonzalez AM (2001) Mitochondrial DNA characterisation of European isolates: the Maragatos from Spain. Eur J Hum Genet 9: 708-716

Lieberman J, Sastre A (1986) Serum angiotensin converting enzyme levels in patients with alpha 1-antitrypsin variants. Am J Med 81: 821-824

Lightowlers RN, Chinnery PF, Turnbull DM, Howell N (1997) Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. Trends Genet 13: 450-455
Lilleväli H, Õunap K, Metspalu A (1996) Phenylalanine hydroxylase gene mutation R408W is present on 84% of Estonian phenylketonuria chromosomes. Eur J Hum Genet 4: 296-300
Loogväli EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk HV, Parik J, Pennarun E, Laos S, Lunkina A, Golubenko M, Barac L, Pericic M, Balanovsky OP, Gusar V, Khusnutdinova EK, Stepanov V, Puzyrev V, Rudan P, Balanovska EV, Grechanina E, Richard C, Moisan JP, Chaventre A, Anagnou NP, Pappa KI, Michalodimitrakis EN, Claustres M, Golge M, Mikerezi I, Usanga E, Villems R (2004) Disuniting Uniformity: A Pied Cladistic Canvas of mtDNA Haplogroup H in Eurasia. Mol Biol Evol 21: 2012-2021
Loze I (1988) Poseleniya kamennogo veka Lubanskoi nizini: Mezolit, ranniy I sredniy neolit. Rīga, Zinātne, 123 lpp.
Loze I (1997) Indo-Europeans in the Eastern Baltic in the view of an archaeologist. Journal of the Institute of Latvian History 3: 16-29
Lum JK, Cann RL (1998) mtDNA and language support a common origin of Micronesians and Polynesians in Island Southeast Asia. Am J Phys Anthropol 105: 109-119
Lutz S, Weisser HJ, Heizmann J, Pollak S (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. Int J Legal Med 111: 67-77
Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2: 13
Macaulay V, Richards M, Sykes B (1997) Mitochondrial DNA recombination – no need to panic. Proc R Soc Lond B Biol Sci 266: 2037-2039
Macaulay VA, Richards MB, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B, Torroni A (1999a) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64: 232-249
Macaulay VA, Richards MB, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B, Torroni A (1999b) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64: 232-249
Maddison D (1991) African origin of human mitochondrial DNA re-examined. Syst Zool 40: 355-363
Maliseva-Lace B, Baumanis V, Krumina A (2004) Clinical heterogeneity of alpha 1-antitrypsin deficiency. Biologija 2: 11-14
Mallory JP (1989) In Search of the Indo-Europeans: Language, Archaeology and Myth. Thames and Hudson, London, pp. 288
Malyarchuk BA, Derenko MV (2001) Mitochondrial DNA variability in Russians and Ukrainians: Implications to the origin of the Eastern Slavs. Ann Hum Genet 65: 63-78
Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Wozniak M, Miscicka-Sliwka D (2002) Mitochondrial DNA variability in Poles and Russians. Ann Hum Genet 66: 261-283
Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Drobnic K, Miscicka-Sliwka D (2003) Mitochondrial DNA variability in Bosnians and Slovenians. Ann Hum Genet 67: 412-425
Manfredi G, Thyagarajn D, Papadopoulou LC, Palloti F, Schon EA (1997) The fate of human sperm-derived mtDNA in somatic cells. Am J Hum Genet 61: 953-960
Marchini J, Cardon LR, Philips MS, Donnelly P (2004) The effects of human population structure on large genetic association studies. Nat Genet 36: 512-517
Margulis L (1975) Symbiotic theory of the origin of eukaryotic organelles; criteria for proff. Symp Sic Exp Biol 29: 21-38
Mason PA, Lightowlers RN (2003) Why do mammalian mitochondria possess a mismatch repair acitivity? FEBS Lett 554: 6-9
McDougall I, Brown FH, Fleagle JG (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature 433: 733–736
McVean GA (2001) What do pattern of genetic variability reveal about mitochondrial recombination? Heredity 87: 613-620
Meinilä M, Finnilä S, Majamaa K (2001) Evidence for mtDNA admixture between the Finns and the Saami. Hum Hered 52: 160-170
Mellars P (2004) Neanderthals and the modern human colonization of Europe. Nature 25: 461-465
Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry SR, Wallace DC (1991) The structure of human mitochondrial DNA variation. J Mol Evol 33: 543-555

Metspalu E, Kivisild T, Kaldma K, Parik J, Reidla M, Tambets K, Villems R (1999) The Trans-Caucasus and the Expansion of the Caucasoid-Specific Human Mitochondrial DNA. In: Papiha S, Deka R, Chakrabotary R (eds) Genomic Diversity: Application in Human Population Genetics. Kluwer Academic/ Plenum Publishers, New York, pp 121-134

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R (2004) Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet 31: 5-26

Meunier J, Eyre-Walker A (2001) The correlation between linkage disequilibrium and distance: implications for recombination in hominid mitochondria. Mol Biol Evol 18: 2132-2135

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA 100: 171-176

Mishmar D, Ruiz-Pesini E, Brandon M, Wallace DC (2004) Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration. Hum Mutat 23: 125-133

Mogentale-Profizi N, Chollet L, Stevanovitch A, Dubut V, Poggi C, Pradie MP, Spadoni JL, Gilles A, Beraud-Colomb E (2001) Mitochondrial DNA sequence diversity in two groups of Italian Veneto speakers from Veneto. Ann Hum Genet 65: 153-166

Molainen JS, Majamaa K (2003) Phylogenetic network and physiochemical properties of nonsynonymous mutations in the protein-coding genes of human mitochondrial DNA. Mol Biol Evol 20: 1195-1210

Mourant AE, Kopec A, Domaniewska-Sobczak K (1976): The Distribution of the Human Blood Groups and Other Polymorphisms. London, Oxford University Press, p. 1055

Mugurevics E (1997) Ethnic progresses in Baltic-inhabited territories and the emergence of the Latvian nation in the 6th to the 16th century. Journal of the Institute of Latvian History 3: 155-167

Nachman MW, Brown WM, Stoneking M, Aquardo CF (1996) Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics 142: 953-963

Nachman MW (1998) Deleterious mutations in animal mitochondrial DNA. Genetics 103: 61-69

Nass MM & Nass S (1963) Intramitochondrial fibers with DNA characteristics. Fixation and electron staining reactions. J Cell Biol 19: 593-611

Nejentsev S, Reijonen H, Adojaan B, Kovalchuk L, Sochnevs A, Schwartz EI, Akerblom HK, Ilonen (1997) The effect of JKL-B allele on the IDDM risk defined by DRB1*04 subtypes and DQB1*0302. Diabetes 46: 1888-1892

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York, pp 145-163

Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, Al-Zahery N, Scozzari R, Cruciani F, Behar DM, Dugoujon JM, Coudray C, Santachiara-Benerecetti AS, Semino O, Bandelt HJ, Torroni A (2006) The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science 314: 1767-1770

Olivo PD, Van de Walle MJ, Laipis PJ, Hauswirth WW (1983) Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. Nature 306: 400-402

Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet 29: 20-21

Oshibkina SV (1983) Mezolit basseina Suhoni i Vistochnogo Prionezhya.Moskva, Nauka, pp. 145

Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, Goodwin W (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature 404: 490-493

Palanichamy M, Sun C, Agrawal S, Bandelt H-J, Kong Q-P, Khan F, Wang C-Y, Chaudhuri T, Palla V, Zhang Y-P (2004) Phylogeny of mtDNA macrohaplogroup N in India based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75: 966-978

Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Allistin-Greiner R, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P, Holland MM (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15: 363-368

Parsons TJ, Irwin JA (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288: 1931

Passarino G, Semino O, Bernini LF, Santachiara-Benecretti AS (1996) Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. Am J Hum Genet 59: 927-934

Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Borresen-Dale AL, Underhill PA (2002) Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. Eur J Hum Genet 10: 521-529

Peltonen L, Palotie A, Lange K (2000) Use of population isolates for mapping complex traits. Nat Rev Genet 1: 182-190
Pereira L, Prata MJ, Amorim A (2000) Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. Ann Hum Genet 64: 491-506
Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A (2001) Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. Ann Hum Genet 65: 439-458
Pereira L, Richards M, Alonso A, Albarran C, Garcia O, Macaulay V, Amorim A (2004) Subdividing mtDNA haplogroup H based on coding-region polymorphisms – a study in Iberia. Int Congr Ser 1261:416-418.
Pesole G, Gissi C, De Chirico A, Saccone C (1999) Nucleotide substitution rate of mammalian mitochondrial genomes. J Mol Evol 48: 427-434
Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P (1999) Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. Int J Legal Med 112: 291-298
Piazza A (1993) Who are the Europeans? Science 260: 1767-1769
Piganeau G, Eyre-Walker A (2004) A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. Heredity 92: 282-288
Poikalainen V (2001) Palaeolithic art from the Danube to Lake Baikal. Folklore 18/19: 7-16
Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc Natl Acad Sci USA 98: 13757-13762
Pronina N, Giannattasio S, Lattanzio P, Lugovska R, Vevere P, Kornejeva A (2003) The molecular basis of phenylketonuria in Latvia. Hum Mutat 21: 398-399
Prīmanis J (1937) Ievads antropoloģijas metodikā. Rīga, Valters un Rapa, 96 lpp.
Pult I, Sajantila A, Simonainen J, Georgiev O, Schaffner W, Pääbo S (1994) Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. Biological Chemistry Hoppe-Seyler 375: 837-840
Puzyrev VP, Stepanov VA, Golubenko MV, Puzyrev KV, Maximova NR, Kharkov VN, Spiridonova MG, Nogovitsyna AN (2003) MtDNA and Y-Chromosome Lineages in the Yakut Population. Genetika 39: 975-981
Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benecretti AS (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat Genet 23: 437-441
Quintana-Murci L, Chaix R, Wells S, Behar D, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Qasim Mehdi S, Torroni A, McElreavey K (2004) Where West meets East: The complex mtDNA landscape of the Southwest and Central Asian corridor. Am J Hum Genet 74: 827-845
Quintans B, Alvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo A (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. Forensic Sci Int 140: 251-257
Rando JC, Pinto F, Gonzalez AM, Nernandez M, Larruga JM, Cabrera VN, Bandelt H-J (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. Ann Hum Genet 62: 531-550
Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, Stoneking M (1995) Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12: 604-615
Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk HV, Parik J, Loogvali EL, Derenko M, Malyarchuk B, Bermisheva M, Zhadanov S, Pennarun E, Gubina M, Golubenko M, Damba L, Fedorova S, Gusar V, Grechanina E, Mikerezi I, Moisan JP, Chaventre A, Khusnutdinova E, Osipova L, Stepanov V, Voevoda M, Achilli A, Rengo C, Rickards O, De Stefano GF, Papiha S, Beckman L, Janicijevic B, Rudan P, Anagnou N, Michalodimitrakis E, Koziel S, Usanga E, Geberhiwot T, Herrnstadt C, Howell N, Torroni A, Villems R (2003) Origin and Diffusion of mtDNA Haplogroup X. Am J Hum Genet 73: 1178-1190
Relethford JH (2001) Ancient DNA and the origin of modern humans. Proc Natl Acad Sci USA 16: 390-391
Renfrew C (1987) Archaeology and Language: the Puzzle of Indo-European Origins. Pimlico, London, pp. 246
Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt H-J, Sykes B (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59: 185-203

Richards MB, Macaulay VA, Bandelt H-J, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62: 241-260

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Golge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Norby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozari R, Torroni A, Bandelt HJ (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67: 1251-1276

Richards M, Macaulay V, Torroni A, Bandelt HJ (2002) In search of geographical patterns in European mitochondrial DNA. Am J Hum Genet 71: 1168-1174

Richards M, Rengo C, Cruciani F, Gratrix F, Wilson J, Scozzari R, Macaulay V, Torroni (2003) Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. Am J Hum Genet 72: 1058-1064

Richards M (2003) The Neolithic invasion of Europe. Ann Rev Anthropol 32: 135-162

Rimantene R (1971) Paleolit i mezolit Litvi. Vilniuss: Mintis, 98 lpp.

Ritte U, Neufeld E, Broit M, Shavit D, Motro U (1993) The differences among Jewish communities – maternal and paternal contributions. I Mol Evol 37: 435-440

Roostalu U, Kutuev I, Loogväli EL, Metspalu E, Tambets K, Reidla M, Khusnutdinova E, Usanga E, Kivisild T, Villems R (2007) Origin and expansion of haplogroup h, the dominant human mitochondrial DNA lineage in West Eurasia: the near eastern and Caucasian perspective. Mol Biol Evol 24: 436-448

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barać L, Peričić M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinnioğlu C, Khusnutdinova E, Rudan P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan JP, Chaventre A, Ferak V, Furedi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzic R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, Villems R, Semino O (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in europe. Am J Hum Genet 75: 128-137

Rosa A, Brehm A, Kivisild T, Metspalu E, Villems R (2004) MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. Ann Hum Genet 68: 340-352.

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

Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. Science 303: 223-226

Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 67: 718-726

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425

Saitou N, Omoto K (1987) Time and place of human origin from mtDNA data. Nature 327: 288

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

Sajantila A, Pääbo S (1995) Language replacement in Scandinavia. Nat Genet 11: 359-360

Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, Macaulay V, Carracedo A (2002) The making of the African mtDNA landscape. Am J Hum Genet 71: 1082-1111

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74: 454-465

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, Second ed. edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 1659

Schatz G, Haslbrunner E, Tuppy H (1964) Deoxyribonucleic acid associated with yeast mitochondria. Biochemical and Biophysical Research Communications 15: 127-132

Schurr TG, Sukernik RI, Starikovskaya YB, Wallace DC (1999) Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea-Bearing Sea region during the Neolithic. Am J Phys Anthropol 108: 1-39

Schurr TG, Wallace DC (2002) Mitochondrial DNA diversity in Southeast Asian populations. Hum Biol 74: 432-452

Schwartz M, Vissing J (2002) Paternal inheritance of mitochondrial DNA. N Egl J Med 347: 576-580

Sedovs V (2004) Balti senatnē. Rīga, Latvijas vēstures institūta apgāds, 166 lpp.

Shadel GS, Clayton DA (1997) Mitochondrial DNA maintenance in vertebrates. Annu Rev Biochem 66: 409-435

Sistonen P, Virtaranta-Knowles K, Denisova R, Kučinskas V, Ambrasiene D, Beckman L (1999) The LWb blood group as a marker of prehistoric Baltic migrations and admixture. Hum Hered 49: 154-158

Smith DG, Malhi RS, Eshleman J, Lorenz JG, Kaestle FA (1999) Distribution of mtDNA haplogroup X among Native North Americans. Am J Phys Anthropol 110: 271-284

Stanevicha V, Eglite H, Sochnevs A, Gardovska D, Zavadska D, Shantere R (2003) HLA class II associations with rheumatic heart disease among clinically homogenous patients in children in Latvia. Arthritis Res Ther 5: 340-346.

Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC (1998) mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of Ancient Beringia and the peopling of the New World. Am J Hum Genet 63: 1473-1491

Stevanovitch A, Gilles A, Bouzaid E, Kefi R, Paris F, Gayraud RP, Spadoni JL, El-Chenawi F, Beraud-Colomb E (2004) Mitochondrial DNA Sequence Diversity in a Sedentary Population from Egypt. Ann Hum Genet 68: 23-29

Stoneking M, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics 124: 717-733

Stringer CB, Andrews P (1988) Genetic and fossil evidence for the origin of modern humans. Science 239: 1263-1268

Stringer C (2003) Human evolution: Out of Ethiopia. Nature 423: 692-3, 695

Sutkovsky O, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G (1999) Ubiquitin tag fro sperm mitochondria. Nature 402: 371-372

Sutkovsky P, Van Leyen K, McCauley T, Day BN, Sutkovsky M (2004) Degradation of paternal mitochondria after fertilization: implications for heteroplasmy, assisted reproductive technologies and mtDNA inheritance. Reprod Biomed Online 8: 24-33

Sykes B, Beiboff A, Low-Beer J, Tetzner S, Richards M (1995) The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. Am J Hum Genet 57: 1463-1475

Tagliabracci A, Turchi C, Buscemi L, Sassaroli C (2001) Polymorphism of the mitochondrial DNA control region in Italians. Int J Legal Med 114: 224-228

Tambets K, Kivisild T, Metspalu E, Parik J, Kaldma K, Laos S, Tolk HV, Gölge M, Demirtas H, Geberhiwot T, De Stefano GF, Papiha SS, Villems R (2000) The topology of the maternal lineages of the Anatolian and Trans-Caucasus populations and the peopling of the Europe: some preliminary considerations. In: Renfrew C, Boyle K (eds) Archaeogenetics: DNA and the population prehistory of Europe: McDonald Institute for Archaeological Research Monograph Series. Cambridge University Press, Cmabridge, pp 219-235

Tambets K, Rootsi S, Kivisild T, Villems R (2001) The concepts of Richard Indreko about the origin of the Finno-Ugric speakers and the population genetics of the extant North-Eastern European populations. TRAMES 5: 59-74

Tambets K, Tolk HV, Kivisild T, Metspalu E, Parik J, Reidla M, Voevoda M, Damba L, Golubenko M, Stepanov V, Puzerev V, Bermisheva M, Knushnudtinova E, Usanga E, Rudan P, Villems R (2003) Complex signals for population expansions in Europe and beyond. In: Bellwood P, Renfrew C (eds) Examining the farming/language dispersal hypothesis. Cambridge University Press, Cambridge, pp 449-458

Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogvali EL, Tolk HV, 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 JP, Barać L, Peričić M, Rudan P, Terzic R, Mikerezi I, Krumina A, Baumanis V, Koziel S, Rickards O, De Stefano GF, Anagnou N, Pappa KI, Michalodimitrakis E, Ferak V, Furedi S, Komel R, Beckman L, Villems R (2004) The Western and Eastern Roots of the Saami--the Story of Genetic "Outliers" Told by Mitochondrial DNA and Y Chromosomes. Am J Hum Genet 74: 661-682

Tanaka M, Cabrera VM, Gonzalez AM, Larruga JM, Takeyasu T, Fuku N, Guo LJ, Hirose R, Fujita Y, Kurata M, Shinoda K, Umetsu K, Yamada Y, Oshida Y, Sato Y, Hattori N, Mizuno Y, Arai Y, Hirose N, Ohta S,

Ogawa O, Tanaka Y, Kawamori R, Shamoto-Nagai M, Maruyama W, Shimokata H, Suzuki R, Shimodaira H (2004) Mitochondrial genome variation in eastern Asia and the peopling of Japan. Genome Res 14: 1832-1850

Templeton A (1992) Human origins and analysis of mitochondrial DNA sequences. Science 255: 737

Templeton A (2002) Out of Africa again and again. Nature 416: 45-51

Thomson G (1991) HLA population genetics. Baillieres Clin Endocrinol Metab 5: 247-260

Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130: 153-162

Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawfoerd MH, Comuzzie AG, Wallace DC (1993a) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet 53: 591-608

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC (1993b) Asian affinities and continental radiation of the four founding Native Amercian mtDNAs. Am J Hum Genet 53: 563-590

Torroni A, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994a) Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North Amercia. Proc Natl Acad Sci USA 91: 1158-1162

Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994b) Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetian population and its adaptation to high altitude. Am J Phys Anthropol 93: 189-199

Torroni A, Chen YS, Semino O, Santachiara-Beneceretti AS, Scott CR, Lott MT, Winter M, Wallace DC (1994c) mtDNA and Y-chromosome polymorphisms in four Native American populations from southern Mexico. Am J Hum Genet 54: 303-318

Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, Wallace DC (1994d) mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. Am J Hum Genet 55: 760-776

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

Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellito D, Rengo C, Forster P, Savontaus ML, Bonne-Tamir B, Scozzari R (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62: 1137-1152

Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R (2001a) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am J Hum Genet 69: 1348-1356

Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E, Parik J, Tolk HV, Tambets K, Forster P, Karger B, Francalacci P, Rudan P, Janicijevic B, Rickards O, Savontaus ML, Huoponen K, Laitinen V, Koivumaki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Benerecetti AS, Semino O, Scozzari R (2001b) A signal, from human mtDNA, of postglacial recolonization in Europe. Am J Hum Genet 69: 844-852

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

van Andel TH (2000) Where received wisdom fails: the Mid-Paleolithic and early Neolithic climates. In: Renfrew C, Boyle K (eds) Archaeogenetics: DNA and the population prehistory of Europe. McDonald institute for Archaeological Research Monograph Series, Cambridge University, Cambridge, pp 31-39

Vasks A (1997) The cultural and ethnic situation in Latvia during the early and middle Iron Age (1st-8th Century AD). Journal of the Institute of Latvian History 3: 168-179

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 86: 9350-9354

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253: 1503-1507

Waeber O (1879) Beitrage zur Anthropologie der Letten. Dissertationes Universitatis Dorpatiensis

Wallace DC (1982) Structure and evolution of organelle genomes. Microbiol Rev 46: 208-240

Wallace DC (1999) Mitochondrial diseases in man and mouse. Science 283: 1482-1488

Wallace DC, Brown MD, Lott MT (1999) Mitochondrial DNA variation in human evolution and disease. Gene 238: 211-230
Wallace DC (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. Annu Rev Genet 39: 359-407
Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61: 691-704
Weidenreich F (1943) The "Neanderthal man" and the ancestors of "Homo sapiens". Am Anthropologist 45: 39-48
Werle E, Schneider C, Renner M, Volker M, Fiehn W (1994) Convenient single-step, one tube purification of PCR products for direct sequencing. Nucleic Acids Res 22: 4354-4355
Viikma M, Heapost L (1996) Genetic differentiation of Estonians. Proc Estonian Acad Sci Biol 45:128–136
White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Howell FC (2003) Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature 12: 742-747
Whittle A (1996) Europe in the Neolithic. Cambridge University Press, Cambridge, pp. 459
Wiik K (2000) European Lingua Francas. In: Künnap A (ed) The Roots of Peoples and Languages of Northern Eurasia II and III. University of Tartu and Societas Historiae Fenno-Ugricae. Tartu, pp 202-236
Wilkinson-Herbots HM, Richards MB, Forster P, Sykes BC (1996) Site 73 in hypervariable region II of the human mitochondrial genome and the origin of European populations. Ann Hum Genet 60: 499-508
Wilson A, Cann R, Carr S, George M, Gyllensten U, Helm-Bychowski K, Higuchi R, Palumbi SR, Prager EM, Sage RD, Stoneking M (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. Biological Journal of the Linnean Society 26: 375-400
Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, Goldstein DB (2001) Genetic evidence for different male and female roles during cultural transitions in the British Isles. Proc Natl Acad Sci USA 98: 5078-5083
Wiuf C (2001) Recombination in human mitochondrial DNA? Genetics 159: 749-756
Wolpoff M, Wu XZ, Thorne AG (1984) Modern Homo sapiens origins: A general theory of hominid evolution involving the fossil evidence from East Asia. In: Smith FH, Spencer F (eds) The origin of modern humans: A world survey of the fossil evidence. Alam L. Riss, New York, pp 411-483
Wolpoff M, Caspari R (1997) Race and Human Evolution. Simon & Schuster, New York, pp. 464
Yao Y, Nie L, Harpending H, Fu Y, Yuan Z, Zhang Y (2002a) genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. Am J Phys Anthropol 118: 63-76
Yao YG, Kong QP, Bandelt H-J, Kivisild T, Zhang YP (2002b) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70: 635-651
Yu MS, Qiu XF, Xue JG, Liu ZD, Tan JZ, Li HJ, Liu DX, Li L, Yu WZ, Tang XZ (1988) Mitochondrial DNA polymorphisms in Chinese. Sci Sin 31: 860-872
Zagorska I (1974) Vēlā paleolīta dzīvesvieta Salaspils Laukskolā. Zinātniskās atskaites sesijas materiāli par arheologu un etnogrāfu 1973. g. pētījumu rezultatiem (A late Paleolithic settlement at the Salaspils rural school: Scientific materials about the results of 1973 research by archaeologists and ethnographists). Rīga, Zinātne
Zagorska I (1981) Agrais mezolīts Latvijas teritorijā (The early Mesolithic period Latvia). LPSR ZA Vēstis, No. 2
Zagorska I (1992) The Mesolithic in Latvia. Acta Archeologica vol. 63
Zagorskis FA (1967) Ranniy i razvitiy neolit v vostochnoy chasti Latvii. Avtoreferat dissertacii na soiskaniye uchennoi stepeni kandidata istoricheskih nauk. Rīga
Zagorskis FA, Zagorska IA (1977) Mezolit Latvii. Kratkiye soobcheniya Instituta Arheologii SSSR, t. 149
Zagorskis FA (1987) Zvejnieku akmens laikmeta kapulauks (The Stone Age Zvejnieki burial ground), Rīga, Zinātne, 124 lpp.
Zaliznyak LL (1979) Ohotniki na severnogo olenya Ukrainskogo Polesya epohi finalynogo paleolita. Kiev: Naukova Dymka
Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60: 1174-1183
Zerjal T, Beckman L, Beckman G, Mikelsaar AV, Krumina A, Kučinskas V, Hurles ME, 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
Zuckerkandl, Emile and Linus Pauling (1962) Molecular disease, evolution, and genetic heterogeneity. In: Horizons in Biochemistry (V. Bryson and H. J. Vogels, eds.), 189-225, Academic Press, New York

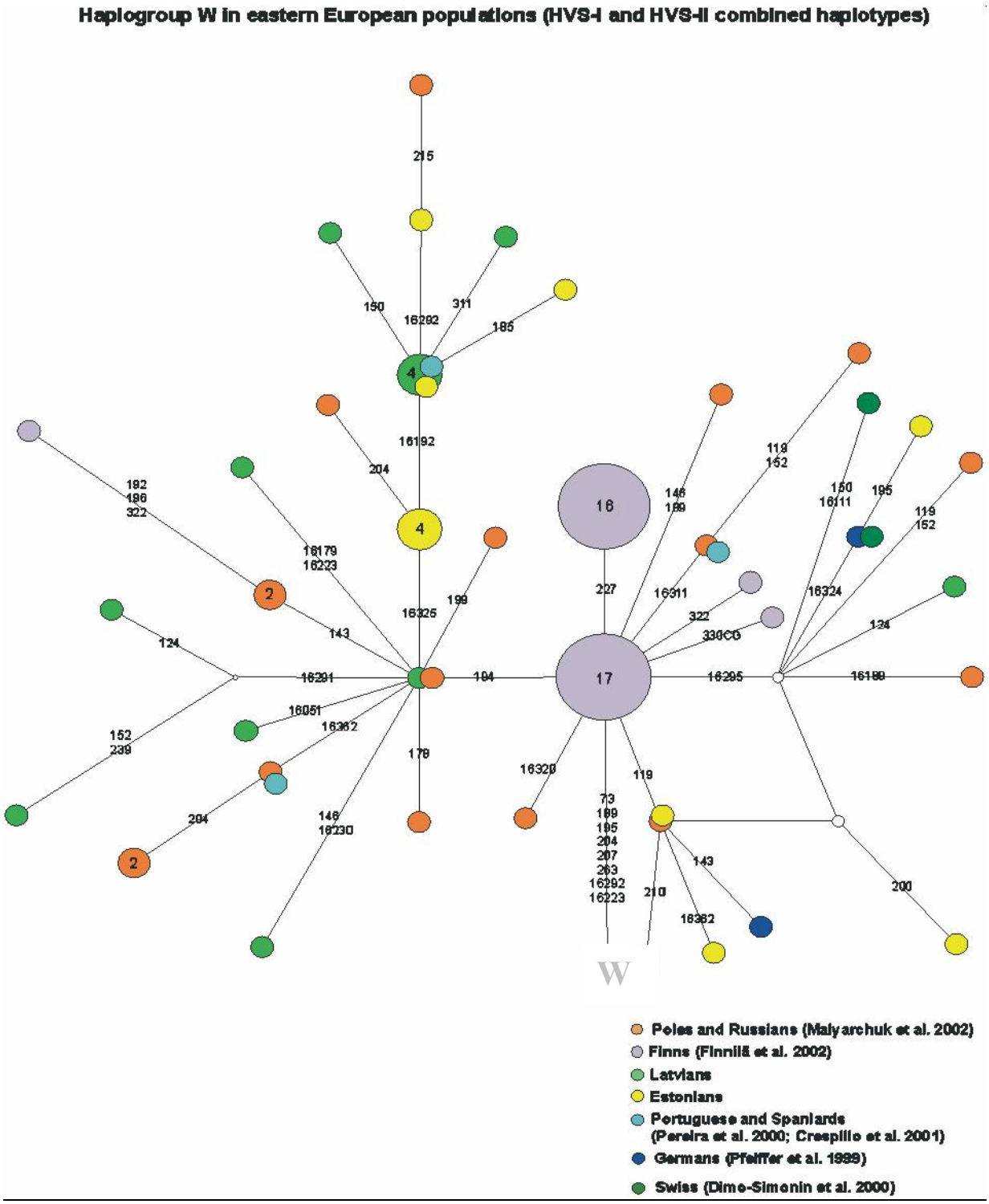
Zvelebil M (1986) Mesolithic prelude and Neolithic revolution. In: Zvelebil M (ed) Hunters in transition: Mesolithic societies of temperate Eurasia and their transition to farming. Cambridge University Press, Cambridge, pp 5-15
Zvelebil M (2000) The social context of the agricultural transition in Europe. In: Renfrew C, Boyle K (eds) Archaeogenetics: DNA and the population prehistory of Europe. McDonald Institute for Archaeological Research Monograph Series, Cambridge University, Cambridge, pp 57-59
Zvelebil M (2001) Revisiting Indreko`s culture historical model: "Origin and area of settlement of the Finno-Ugrian peoples". TRAMES 5: 26-47

Electronic-Database Information

The URL for data presented herein is as follows:
Fluxus Engineering, http://www.fluxus-engineering.com/
MITOMAP: A Human Mitochondrial Genome Database, http://www.mitomap.org, 2006

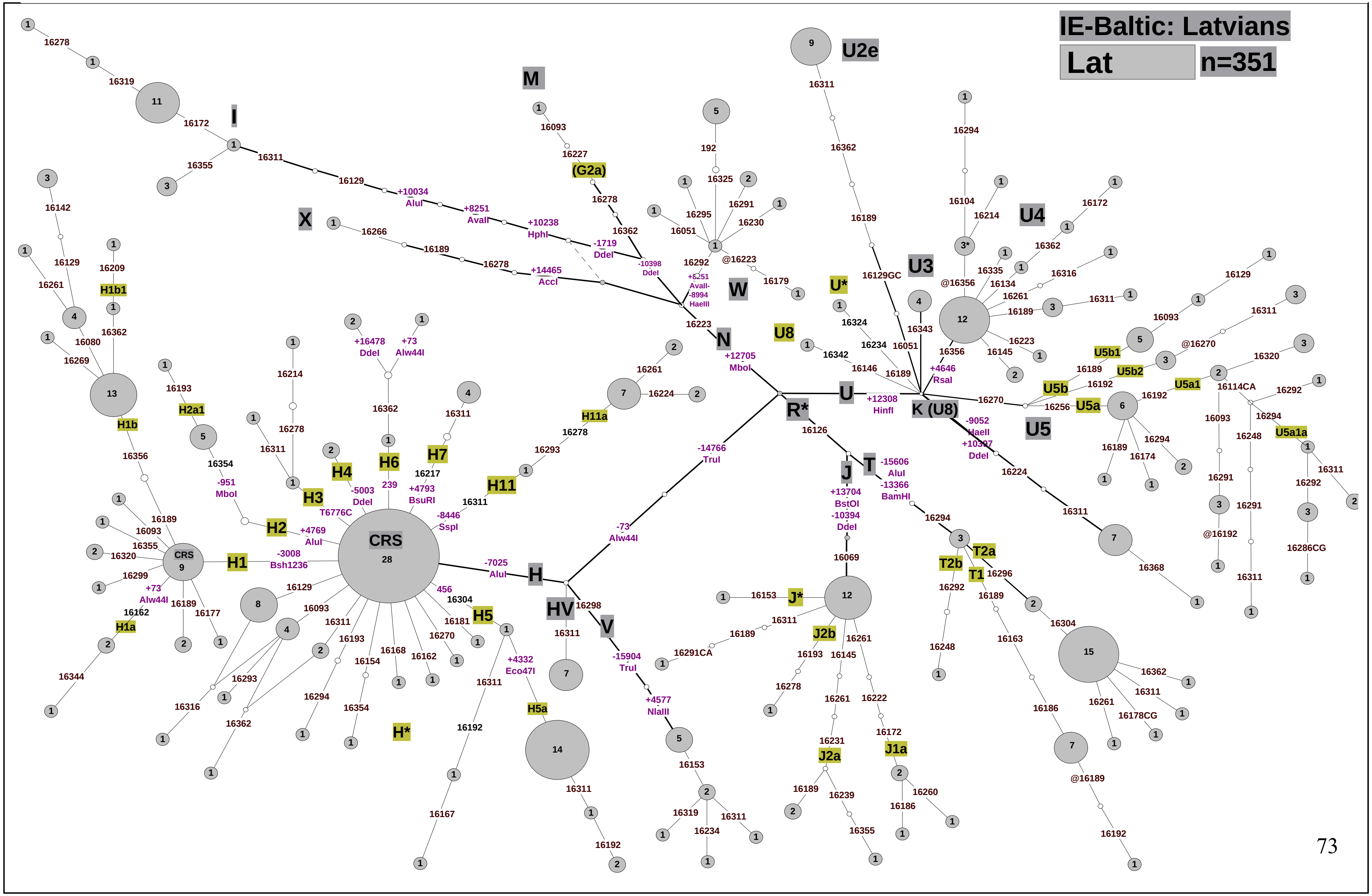
APPENDIX

Appendix 1. The phylogenetic network of mtDNA haplogroup W haplotypes found among Latvians and several European populations



Numbers on the links indicate observed mutations and are numbered according to the revised Cambridge Reference Sequence (Andrews *et al.*, 1999); nucleotide change is specified by suffixes only for transversions. Numbers on the links indicate observed mutations and are numbered according to the revised Cambridge Reference Sequence (Andrews *et al.*, 1999); nucleotide change is specified by suffixes only for transversions.

Appendix 2. The phylogenetic network of mtDNA haplotypes found among Latvians. Numbers on the links indicate observed mutations and are numbered according to the revised Cambridge Reference Sequence (Andrews *et al.*, 1999); nucleotide change is specified by suffixes only for transversions. The gain or loss of a restriction site is marked by a "+" or a "-", respectively; symbol @ denotes a recurrent mutation.



AKNOWLEGMENTS

This study was carried out in the Latvian Biomedical Research and Study Centre and at the Department of Evolutionary Biology, Institute of Molecular and Cell Biology of the University of Tartu.
My deepest gratitude goes to my supervisor Prof. Viesturs Baumanis, who has been the main driving force in formulating the most interesting ideas, questions to investigate, and for his guidance through the hills and valleys of the scientific world. His help has been inestimable for the completion of this work.
This work became essentially possible through the collaboration with several people from different Institutions. I would especially like to thank Prof. Astrida Krumina, for the priceless collection of DNA samples, for her invaluable assistance in all stages of work, Prof. Richard Villems, for his help in writing the articles, for introduction to population genetics of uniparentally inherited markers and for offering an opportunity to perform the greatest part of the experimental work in his laboratory. I have great pleasure to thank Prof. Laurent Excoffier, for his assistance with statistics. I would like to give my special thanks to Kristiina Tambets for her permanent patience, cheerful attitude and practical support. I am also indebted to Prof. Arturs Socnevs for the Lettigalian collection.
I would like to express my sincere thanks to all the colleagues and fellow students (especially to Liga Timsa) from the Department of Molecular Microbiology in the Latvian Biomedical Research and Study Centre, for creating an inspiring and pleasant atmosphere in the laboratory, for interesting discussions and valuable criticism during the years of my research work.
I place great value in the friendly environment and people of our Institute, Riga Stradins University (Department of Medical Biology and Genetics) and Institute of Molecular and Cell Biology of the University of Tartu (Department of Evolutionary Biology). I appreciate the excellent technical assistance of Eva Stankevica, Ligija Ignatovica, Ille Hilpus and Jaan Lind. And then, I owe a lot to my dearest one, from whom I inherited and share with the identical mtDNA variant. She has encouraged and supported me constantly during my studies.
This research was supported by a grant from the European Social Fund (ESF) program ESS2004/3 and Latvian Council of Science National program 01.0023.

ORIGINAL PUBLICATIONS