DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
102

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
102

HUMAN Y-CHROMOSOMAL VARIATION IN EUROPEAN POPULATIONS

SIIRI ROOTSI

TARTU UNIVERSITY
PRESS

Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (in molecular biology) on October 19, 2004 by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Opponent: Dr. Peter Forster, Molecular Genetics Laboratory, McDonald Institute for Archaeological Research, Cambridge University, UK

Commencement: Room No 218, Riia 23, Tartu, on Dec. 10 at 12.15

The publication of this dissertation is granted by the University of Tartu

Tartu Ülikooli Kirjastus
www.tyk.ut.ee
Tellimus nr. 547

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

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

I. **Rootsi, S.**, Magri, C., Kivisild, K., Benuzzi, G., Help, H., Bermisheva, M., Kutuev, I., Barać, L., Peričić, M., Balanovsky, O., Pshenichnov, A., Dion, D., Grobei, M., Zhivotovsky, L. A., 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, J.-P., Chaventre, A., Ferak, V., Füredi, S., Oefner, P. O., Shen, P., Beckman, L., Mikerezi, I., Terzić, R., Primorac, D., Cambon-Thomsen, A., Krumina, A., Torroni, A., Underhill, P. A., Santachiara-Benerecetti, A. S., Villems, R., Semino, O. (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *American Journal of Human Genetics* 75, 128–137.

II. Tambets, K**., Rootsi, S**., Kivisild, T., Help, H., Serk, P., Loogväli, E.-L., Tolk H.-V., Reidla, M., Metspalu, E., Pliss, L., Balanovsky, O., Pshenichnov, A., Balanovska, E., Gubina, M., Zhadanov, S., Osipova, L., Damba, L., Voevoda, M., Kutuev, I., Bermisheva, M., Khusnutdinova E., Gusar, V., Grechanina, E., Parik, J., Pennarun, E., Richard, C., Chaventre, A., Moisan, J.-P., Barać, L., Peričić, M., Rudan, P., Terzić, R., Mikerezi, I., Krumina, A., Baumanis, V., Koziel, S., Rickards, O., De Stefano, GF., Anagnou, N., Pappa, K.I., Michalodimitrakis, E., Ferák, V., Füredi, S., Komel, R., Beckman, L., Villems, R. (2004) The western and eastern roots of the Saami — the story of genetic "outliers" told by mtDNA and Y-chromosome. *American Journal of Human Genetics* 74, 661–682.

III. Barać L, Peričić M, Martinović Klaric I, **Rootsi S,** Janicijević B, Kivisild T, Parik J, Rudan I, Villems R, Rudan P (2003) Y chromosomal heritage of Croatian population and its island isolates. *Eur J Hum Genet*. 11, 535–542.

IV. 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-East European populations. *TRAMES*, 5 (55/50), 1, 59–74.

Original publications are reproduced with permission from the publishers.

The author's contribution to the articles referred in the current thesis is as follows:

Ref. I — conceived and designed the experiments, performed the experiments for populations listed in Subjects and Methods section, analyzed the data, wrote the paper;

Ref. II, IV — conceived and designed the Y-chromosomal experiments, performed the experiments of the Y-chromosomal part of study as listed in Subjects and Methods, performed the phylogeographical analysis of Y-chromosomal data, wrote the paper;

Ref. III — assisted in performing the experiments; participated in the analysis of the data and writing the manuscript

ABBREVIATIONS

AZF	azoospermia factor
Bp, kbp, Mbp	base pair, thousand base pairs, million base pairs
DAZ	Deleted in AZoospermia
DHPLC	denaturing high performance liquid chromatography
DNA	deoxyribonucleic acid
Hg	haplogroup
IR	Inverted Repeat
Kya	thousand years ago
LGM	Last Glacial Maximum
MSY	male-specific region of Y chromosome
MtDNA	mitochondrial DNA
NRPY	non-recombining portion of Y
NRY	non-recombining region of Y chromosome
PCR	polymerase chain reaction
RFLP	Restriction Fragment Length Polymorphism
SNP	single nucleotide polymorphism
SRY	Sex determing region Y
STR	short tandem repeat
TMRCA	time to most recent common ancestor
VNTRs	variable number of tandem repeats
YAP	ALU polymorphism in Y chromosome
YCC	Y Chromosome Consortium
Yp	short arm of Y chromosome
Yq	long arm of Y chromosome

1. INTRODUCTION

The main biological importance of Y chromosome is its role in sex determination and male fertility. Understanding its genetics is, therefore, of a wide medical importance. That, however, does not exhaust its use as an object of research. Evolution of sex is among the basic problems of evolution. Furthermore, as we have witnessed recently, the Y chromosome became a powerful instrument to study population genetics of bisexual organisms, including mammals. As far as humans are concerned, studying Y-chromosomal variation has lasted nearly 20 years by now.

For a long time the Y chromosome was considered to be quite non-polymorphic chromosome, consisting of lots of junk-DNA and containing very few genes. Only in the last decade and in particular during the last five years**,** many new studies of Y chromosome have considerably enhanced our knowledge about different aspects of the structure and function of this haploid genome. By now, the euchromatic region of the human Y chromosome has been completely sequenced, revealing different functional regions and containing more than 20 genes. Many polymorphisms (including more than 300 biallelic markers) in the non-recombining region of Y (NRY) have been described up to now, constantly improving the resolution of the phylogenetic tree of the Y chromosome.

Its specific features as haploid state, the presence of large non-recombining areas and patrilinear transmission, make Y-chromosome a unique and powerful research instrument for many different fields of biomedical sciences, including population genetics and evolutionary studies in general, in applied fields like forensic studies etc.

The first part of the text gives an overview of general aspects of Y chromosome structure and features that make it useful for studying phylogeny and phylogeographical spread of Y-chromosomal lineages worldwide.

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 Y-chromosomal variation, in particular as far as European populations are concerned.

2. LITERATURE OVERVIEW

2.1. Structure of human Y chromosome

Y chromosome contains about 60 million base pairs (Mbp) of DNA. 95% of the length of Y chromosome encompasses region, where there is no X–Y crossing-over in male meiosis and it is called the non-recombining region of Y (NRY), non-recombining portion Y (NRPY) or the male-specific region (MSY). The NRY is flanked on both sides by pseudoautosomal regions, where X–Y crossing over is a normal and frequent event in male meiosis (Simmler et al. 1985; Cooke et al. 1985; Freije et al. 1992).

A detailed physical map of human Y chromosome was obtained by Tilford et al. (2001). The first reports about the nucleotide sequence of two portions of the NRY (AZFa and AZFc regions) were published by Sun et al. (1999) and Kuroda-Kawaguchi et al. (2001). These results were recently incorporated to the analysis of the entire NRY nucleotide sequence (Skaletsky et al. 2003). 97% of the NRY from one man was sequenced and it was found to contain at least 156 transcription units, all located within its euchromatic sequences. Half of the transcription units encode 27 distinct proteins or protein families, 12 of which are expressed ubiquitously in most tissues while 11 are testes-specific. These findings confirm a previous model, proposing two distinct functional classes of NRY genes (Lahn and Page 1997).

2.1.1. Non-recombining region of the Y chromosome (NRY)

NRY of the Y chromosome (fig. 1) splits roughly into two large parts: euchromatic and heterochromatic portions.

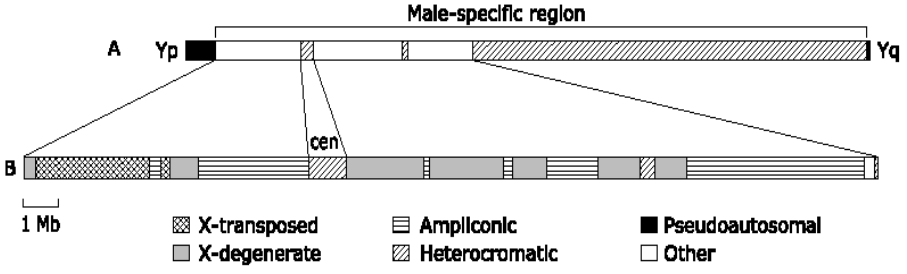

Figure 1. Male-specific region of the Y chromosome (adapted from Skaletsky et al. 2003). A: Schematic representation of the whole chromosome, including the pseudoautosomal and heterochromatic regions.
B: Enlarged view of a 24-Mb portion of euchromatic region of the NRY. Different euchromatic sequence classes are shown. A 1-Mb bar indicates the scale of diagram.

The satellite sequences were equated to heterochromatin and all other sequences with euchromatin (Skaletsky et al. 2003).

2.1.1.1. Heterochromatic regions of NRY

Efforts to gain sequence-based understanding of human chromosomes have largely by-passed heterochromatic regions (Venter et al. 2001). In addition to earlier known centromeric heterochromatin (Tyler-Smith et al. 1993) and much longer heterochromatic block (roughly 40 Mb) that comprises the bulk of the distal long arm, a third heterochromatic block — a sharply demarcated island that spans approximately 400 kb, comprises 3,000 tandem repeats of 125 base pairs (bp) and interrupts the euchromatic sequences of proximal long arm of Y chromosome (Yq) — was discovered and characterized by Skaletsky et al. (2003), see fig. 1. It was found, that the heterochromatin of NRY encompasses at least six distinct sequence classes, each of which form long, homogeneous tandem arrays (Skaletsky et al. 2003).

2.1.1.2. Euchromatic region of NRY and its characteristic sequence classes

The euchromatic DNA sequences total roughly 24 Mb, including 8Mb on the short arm (Yp) and 14.5Mb on the long arm (Yq), with two minor gaps (fig. 1B). Nearly all of the euchromatic sequences fall into three distinctive classes — X-transposed, X-degenerated and ampliconic segments.

The presence of X-transposed sequences in the human MSY is a result of a massive X-to-Y transposition that occurred about 3–4 million years ago, after the divergence of the human and chimpanzee lineages (Rozen et al. 2003) and they are 99% identical to DNA sequences on the long arm of X chromosome (Xq21). Subsequently, an inversion within the NRY short arm cleaved the X-transposed block into two non-contiguous segments, as observed in the modern human NRY (fig. 1B). The X-transposed sequences do not participate in X–Y crossing over during male meiosis, distinguishing them from the pseudoautosomal sequences. Within the X-transposed segments, which have a combined length of 3.4Mb, only two genes were identified, both of which have homologues on Xq21. Thus, the X-transposed sequences exhibit the lowest density of genes among the three sequence classes in the NRY euchromatin, as well as the highest density of interspersed repeat elements (Skaletsky et al. 2003).

In contrast to X-transposed regions, the X-degenerate segments of the NRY are dotted with single-copy gene or pseudogene homologues of 27 different X-linked genes. These single-copy NRY genes and pseudogenes display between 60% and 96% nucleotide sequence identity to their X-linked homologues, and they seem to be surviving relics of ancient autosomes from which the X and Y

chromosomes co-evolved (Lahn and Page 1999). In 13 cases, the NRY homologue is a pseudogene, in the remaining cases, the NRY homologue seems to be a transcribed, functional gene, and the X- and Y-linked genes encode very similar but non-identical protein isoforms. Notably, all 12 ubiquitously expressed NRY genes reside in the X-degenerate regions. Conversely, among the 11 NRY genes expressed predominantly in testes, only one gene, the sex-determining region (SRY), is X-degenerate (Skaletsky et al. 2003).

The third class, ampliconic segments or segmentally duplicated portion of NRY- name introduced by Hurles and Jobling (2003) — are composed largely of sequences that exhibit marked similarity: as much as 99.9% identity to other sequences in the NRY (Skaletsky et al. 2003). The amplicons are located in seven segments that are scattered across the euchromatic long arm and the proximal part of the short arm (fig. 1B) with their combined length of 10.2Mb. The ampliconic sequences exhibit the highest density of genes among the three sequence classes in the NRY euchromatin. Nine distinct NRY-specific protein-coding gene families were identified, with copy numbers ranging from two to approximately 35. All protein-coding families in the ampliconic regions are expressed predominantly or exclusively in testes.

The most pronounced structural features of the ampliconic regions are eight massive palindromes, with arm-to-arm nucleotide identities of 99.94–99.997%. The eight palindromes collectively comprise 5.7Mb, or one quarter of the NRY euchromatin. Six of the eight palindromes carry recognized protein-coding genes, all of which seem to be expressed specifically in testes (Skaletsky et al. 2003). In all known cases of genes on NRY palindromes, identical or nearly identical gene copies exist on opposite arms of the palindrome.

In addition to the palindromes, the ampliconic regions contain five sets of more widely spaced inverted repeats (IRs). Three of these (IR1, IR2 and IR3) exhibit nucleotide identities of 99.66–99.95%. The ampliconic regions contain also a variety of long tandem arrays.

2.2. Evolution of the Y chromosome

The mammalian X and Y chromosomes are thought to have evolved from an ordinary pair of autosomes (Ohno 1967; Graves and Schmidt 1992). Support for this hypothesis, and a proposed 300-million-year timeline for sex chromosome evolution, have emerged from studies of modern X–Y gene pairs (Jegalian and Page 1998; Lahn and Page 1999). Lahn and Page (1999) reasoned that X–Y differentiation would have begun only after X–Y crossing over ceased and the first event, which marked the beginnings of X-Y differentiation, occurred about 240 to 320 million years ago, shortly after divergence of the mammalian and avian lineages. Among the 19 X–Y gene pairs studied, age increased in a stepwise fashion along the length of the X chromosome, in four ‘evolutionary

strata'. This suggested that at least four events had punctuated sex chromosome evolution, with each event suppressing X–Y crossing over in one stratum without grossly disturbing gene order in the X chromosome. Compared to previous estimates the results of Skaletsky et al. (2003) and Rozen et al. (2003) give the time scale that extends from approximately 4 million years for the X-transposed sequences — the youngest known sequences in the NRY, to approximately 300 million years for SRY — the sex determinant and arguably the oldest gene in the NRY.

All NRY X-degenerate genes and pseudogenes seem to be products of a single molecular evolutionary process: the region-by-region suppression of crossing over in ancestral autosomes, with subsequent differentiation of the Y from the X chromosome (Charlesworth 1996; Graves 1996; Lahn and Page 1999). At least two of the NRY's ampliconic gene families are thought to be also originated in this manner, but subsequently acquired the characteristics of ampliconic sequences. Inversions in the Y chromosome may have suppressed crossing over with the X chromosome. The findings about the X-degenerate regions support the theory of the genetic benefits of sexual recombination through meiotic crossing over, and the deleterious consequences of its absence (Skaletsky et al. 2003). According to this theory, most ancestral genes remained functionally intact in the X chromosome, where the benefits of crossing over (in females) continued. In the Y chromosome, in contrast, the shutting down of X–Y crossing over during evolution triggered a monotonic decline in gene function. In this light, the protein-coding genes in the modern NRY's X-degenerate sequences appear as rare examples of persistence in the absence of sexual recombination.

The situation is different for ampliconic genes. Despite the wide variety of genomic sources and mechanisms that gave rise to the ampliconic genes (Skaletsky et al. 2003), they came to exist in the NRY in multiple, nearly identical copies in palindromes. Therefore, the understanding, how the ampliconic genes avoid degradation due to mutation in the absence of crossing-over may be an important outcome from the sequencing of NRY. The possible mechanism that preserves the genes in palindromes is likely to be gene conversion by which the Y chromosome repairs mutations that occur within these genes (Rozen et al. 2003). The occurrence of NRY gene pairs that are subject to frequent gene conversion might provide a mechanism for conserving gene functions across evolutionary time in the absence of crossing-over.

2.3. Y chromosome as a tool to study human genetic variation and demographic history

2.3.1. Special features of Y chromosome

The reasons that make Y chromosome a suitable tool for investigating the recent human evolution (Jobling and Tyler-Smith 1995; Underhill et al. 2000;

2003 Hammer et al. 2001; Hammer and Zegura 2002), for medical genetics (Jobling and Tyler-Smith 2000), DNA forensics (Jobling et al. 1997) and genealogical reconstructions (Jobling 2001), result from its uniqueness among the other human chromosomes. The Y chromosome has a sex-determing role, it is male specific and constitutively haploid. It is inherited paternally and is transmitted from father to son, and unlike other chromosomes, the Y chromosome escapes meiotic recombination in its NRY region. The main importance of the lack of recombination is that haplotypes that form on the basis of the combinations of allelic states of markers usually pass intact from generation to generation. In other words, the non-recombining portion of the Y chromosome descends as a single locus. As they change only by accumulating mutations in time, they preserve by far more simple record of their history compared to autosomes.

Y chromosome has an additional specific feature due to its singularity. Namely, assuming an 1:1 sex ratio, the effective population size of Y in whole population is expected to be one-quarter of that of any autosome, one third of X-chromosome and similar to that of mtDNA. Accordingly, the Y chromosomal genetic variation is, compared to that of autosomes, more susceptible to random genetic drift that modifies the frequencies of different haplotypes, particularly in small populations.

2.3.2. Types of Y-chromosomal markers

Y chromosome variation consists of large amount of different types of polymorphisms, which are widely used in evolutionary studies. They may roughly be divided into two large groups: bi-allelic markers and polymorphisms of tandem repeats or multi-allelic markers (Jobling and Tyler-Smith 2000; Hammer and Zegura 2002).

Biallelic markers include SNPs (Single Nucleotide Polymorphisms) and insertions and deletions (indels). SNPs are the most common type of polymorphisms, constituting more than 90% of total polymorhpisms of DNA (Collins et al. 1997). Only these bi-allelic mutations that have occurred, highly likely, only once in history of humans and have a detectable frequency in human populations are used in phylogenetic studies. Sometimes recurrent mutations are also used (YCC 2002), but then the recurrent mutation has to have occurred in different, unambiguously independently distinguishable branches in phylogenetic tree of Y chromosome, like SRY 1532 in background of haplogroups A and R1a.

SNPs are characterized by low mutation rate and are therefore suitable for studying early demographic events in human history. The mutation rate for SNP markers is considered to be an average on the order of 2×10^{-8} per base per generation (Nachman and Crowell 2000). In the case of indels (YAP insertion, 12f2 deletion), their presence or absence compared to ancestral state is detected.

Large rearrangements, mostly deletions in regions of Y-specific genes (AZFa, AZFb, AZFc), have been known as causes for many diseases leading to male infertility, causing spermatogenic failure, azoospermia, severe oligospermia or otherwise severely impair male reproductive fitness (Vogt et al. 1996; 1998; Yen et al. 1998; Blanco et al. 2000; Krausz et al. 2000; Fernandes et al. 2002; 2004; Repping et al. 2003; 2004).

Not all indels affect male fertility: they persist over generations and are sufficiently common to be considered as polymorphisms. One such example is a 2kb deletion in 12f2 marker (Casanova et al. 1985), used for defining haplogroup J according to the present nomenclature (YCC 2002). Some indels have arisen independently more than once in human history. For example, the deletion or duplication of the 50f2/C (DYS7C) region in background of different haplogroups is thought to be arisen at least 7–8 times (Jobling et al. 1996). Another example is the deletion of DAZ3/DAZ4 region that has been indicated to occur in haplogroup N individuals (Fernandes et al. 2004), widely spread in northern Eurasia. These findings show that new informative variations in different Y-chromosomal haplogroup backgrounds in phylogenetic tree may be detected also in studies, otherwise focused on medical aspects.

Another frequent type of polymorphisms, present also in Y chromosome, is tandem repeats, mostly in non-coding DNA regions (for a recent review, see Chambers and MacAvoy 2000). According to their length, these repeats are classified as satellite-DNAs (repeat lengths of one to several thousand base pairs), minisatellites or variable number of tandem repeats, (VNTRs) ranging from 10 to 100 bp, and microsatellites or short tandem repeats (STRs), with motifs less than 10 bp, mostly 2 to 6 bp long (Nakamura et al. 1987; Charlesworth et al. 1994; Chambers and MacAvoy 2000).

In Y-chromosomal studies microsatellites are widely used, while minisatellites have been used only in some investigations (e.g. Jobling et al. 1998; Bao et al. 2000; Jin et al. 2003).

Microsatellites are multi-allelic markers with different allele numbers ranging from 3 to 49 in locus (de Knijff et al. 1997) and were first taken in use by Litt and Luty (1989). Their mutation rate is much higher than that for bi-allelic markers and, therefore, they are widely used in phylogenetic studies to investigate details of demographic events that have occurred in a more recent time-scale. In evolutionary studies STRs are valuable in combination with binary haplogroup data (de Knijff 2000), as they enable to study diversity within a haplogroup. STRs are particularly widely explored in forensic work (Jobling et al. 1997). So far the number of widely used Y-chromosomal STRs has been quite low (about 30) but in a recent study by Kayser et al. (2004), 166 new and potentially useful STRs were described.

Different analyses have shown that the average mutation rate for autosomal tetranucleotide repeats is about 2.0×10^{-3} per generation (Weber and Wong 1993). Similar results (2.0×10^{-3} per generation) were obtained also for Y-

chromosomal tetranucleotide repeats in deep pedigree studies (Heyer et al. 1997) that were in concordance with results of Weber and Wong (1993) for autosomal microsatellites. Mutation rate was studied in father-son pairs, getting average mutation rate 2.8 x 10^{-3} by Kayser et al. (2000b), while in Forster et al. (2000), by using the haplotype network of Native Americans the average rate was found to be 2.6 x 10^{-4} / per generation (here: 20 years) per locus, differing thus about an order of magnitude from results by Heyer et al. (1997). In Zhivotovky et al. (2004), mutation rate was estimated, using data on microsatellite variation within Y chromosome haplogroups defined by SNPs in populations with documented short-term histories (the African Bantu expansion, the divergence of Polynesian populations and the origin of Gypsy populations from Bulgaria), as well as making use of comparative data on worldwide SNP variation, both at autosomal and Y chromosome loci. The estimated mutation rate for an average Y chromosome short-tandem repeat locus was found to be 6.9 x 10^{-4} per generation (here: 25 years).

2.4. Phylogeny of Y-chromosomal haplogroups

First studies in this field were initiated in mid-1980s, when the first polymorphisms in the Y chromosome were reported: 12f2 marker (Casanova et al. 1985) and 49a,f polymorphism (Ngo et al. 1986) which molecular basis was described by Jovelin et al. (2003). At the beginning, the Y chromosome was thought to be extremely uniform and non-polymorphic (reviewed in Jobling and Tyler-Smith 1995). Together with improvement of new technologies and methods (PCR, large-scale sequencing projects, DHPLC), the number of markers started to increase. Until 1997, there were only 11 known binary polymorphisms that could be genotyped by PCR-based methods (Jobling et al. 1997). Thereafter, Underhill et al. (1997) published 19 new PCR-based binary markers that were discovered and detected by denaturing high performance liquid chromatography (DHPLC). Since then, this method has been used to discover more than 300 SNPs and small insertions/deletions on the NRY (Shen et al. 2000; Underhill et al. 2000; Hammer et al. 2001; Underhill 2003).

Many groups started to screen populations from different regions in large scales for various binary Y chromosomal polymophisms. At least seven different nomenclatures existed (Su et al. 1999; Jobling and Tyler-Smith 2000 and Kalaydjieva et al. 2000; Semino et al. 2000; Underhill et al. 2000; Capelli et al. 2001; Hammer et al. 2001; Karafet et al. 2001), whereas haplogroups/types in them were defined by different, only partially overlapping sets of markers. Consequently, every nomenclature had its unique symbols to label Y chromosome haplogroups/types.

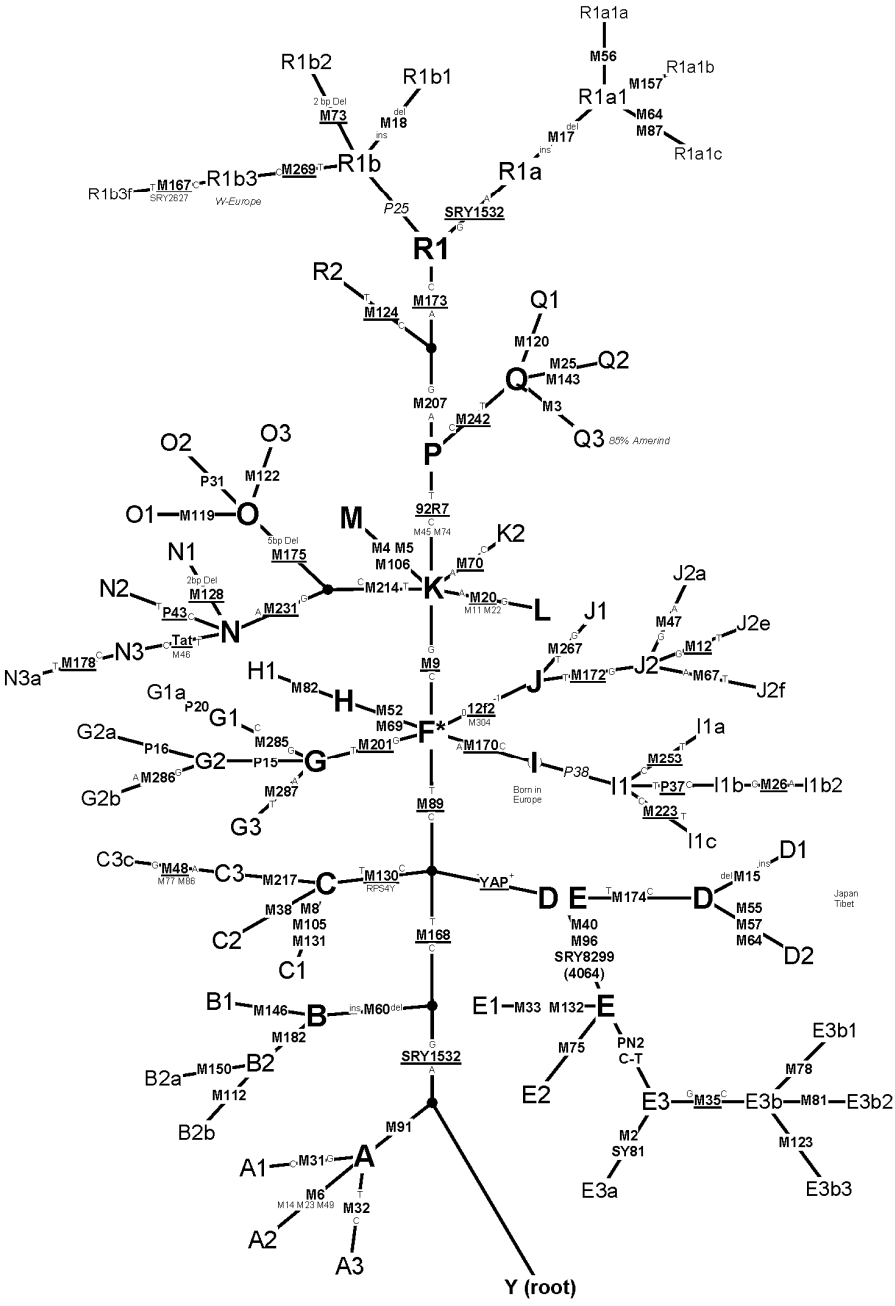

Figure 2 Schematic most parsimonious phylogenetic tree of Y-chromosomal haplogroups, re-designed here from Phylogenetic Tree of YCC (Jobling and Tyler-Smith 2003). Haplogroups are indicated with capital letters and markers defining them are shown on lines.

It became complicated to follow and compare results obtained and published by different groups. From this practical necessity, the Y Chromosome Consortium (YCC) has by now developed a synthetic (synthetic in this context — as a synthesis of virtually all nomenclatures existing at early 2000's), most parsimonious Y-chromosomal phylogenetic tree of binary haplogroups (YCC 2002, see fig. 2). The tree was rooted by outgroup comparisons. That means, whenever possible, homologous regions of the NRY in closely related species (chimpanzees, gorillas and orangutans) were sequenced to determine the likely ancestral states for the established in human Y chromosome polymorphic sites (Underhill et al. 2000; Hammer et al. 2001).The new nomenclature (YCC 2002) included 245 binary markers that were genotyped in globally representative set of samples. Altogether 153 binary haplogroups were observed and a single, most parsimonious phylogeny was constructed.

Term haplogroup refers to NRY lineages defined by binary polymorphisms (YCC). Defining the monophyletic haplogroups in phylogenetic tree is based on derived states of biallelic markers. Lineages that are not defined on the basis of derived states of markers represent interior nodes of the haplogroup tree and are potentially paraphyletic — representing chromosomes that belong to a clade but not its sub-clades — named paragroups by YCC (2002) and distinguished from haplogroups by * symbol. Term *haplotype* is reserved for groups of Y chromosomes defined by STRs variation (de Knjiff 2000; YCC 2002).

The advantages of the nomenclature are: a) haplogroups are placed on tree in hierarchical order; b) flexibility in naming of haplogroups, as well as standardizing the earlier used names; c) ability to accommodate new haplogroups as new mutations are discovered and annually republished**,** reflecting the changes resulting from new discoveries in this field. By now, the second edition of YCC nomenclature is published (Jobling and Tyler-Smith 2003) with some minor changes and refinements compared to the original one. The new nomenclature enables much easier comparison between different datasets. Nevertheless, because of lower phylogenetic resolution, data from earlier publications pose several problems. Unless they were updated to present resolution, the value of such data sets is going to diminish. In the present thesis, the haplogroups are named according to the nomenclature as it was proposed in YCC (2002; 2003) even when referring to papers published before the common nomenclature. In some (controversial) cases additional explanations are given.

2.5. Recent African origin of anatomically modern humans and factors influencing their spread

Recent science history knows two opposite theories about when and where the initial colonization of the World by modern humans has started: (a) the "recent Out-of-Africa" (e.g. Cann et al. 1987; Stringer and Andrews 1988) and (b) the

multiregional evolution model, the latter propagated currently most vocally by Wolpoff (Wolpoff et al. 1988; 1989; Wolpoff et al. 2001). The first widely publicized evidence supporting "recent Out of Africa" colonization theory came from studies of human molecular diversity at the end of 1980s and beginning of 1990s (Wainscoat et al. 1986; Cann et al. 1987; Vigilant et al. 1991). These studies suggested that our species had evolved from a relatively small African population that had subsequently colonized the whole World, supplanting former hominoids. Many studies have shown that African populations harbour more genetic diversity than non-African populations in mtDNA (Cann et al. 1987; Vigilant et al. 1991; Ingman et al. 2000), in Y chromosome microsatellites and biallelic markers (Seielstad et al. 1999; Hammer et al. 2001; Underhill et al. 2001a) and in autosomal STRs and SNPs (Calafell et al. 1998).This theory is also supported by an evidence that allelic diversity outside of Africa is often essentially a subset of that found within Africa (Armour et al. 1996; Calafell et al. 1998; Kivisild et al. 1999; Yu et al. 2002). Recent "Out of Africa" has gained support and independently suggested using archaeological evidence (Stringer and Andrews 1988; Stringer and McKie 1996; Stringer 2003), while a recent discovery of Homo sapiens fossils in Ethiopia dating to 160 000 years ago is a further argument in favor of a recent African origin of our species (White et al. 2003).

By now, the "Out of Africa" model has been overwhelmingly accepted by geneticists and the current "center of gravity" of research and debating has rather shifted to the possible migratory routes of the African exodus and its time-scale (e.g. Lahr and Foley 1998; Stringer 2000; Quintana-Murci et al. 1999; Kivisild et al. 1999; 2003; Underhill et al. 2001a; Underhill 2003; Cavalli-Sforza and Feldman 2003; Metspalu et al. 2004).

Phylogeography is the analysis of the geographical distribution of the different branches (genealogical lineages) of phylogenetic tree (Avise et al. 1987; Avise 2000).

The extant phylogeographic spread of Y-chromosomal variation has been influenced by many past events that have shaped demographic history of populations and occurred over a long prehistoric and historic span of time during the existence of our species. Traces of such events are usually reshaped many times. Therefore, data provided by other genetic markers, as well as additional information that can be obtained from classical paleoanthropology, including archaeology, understanding dynamics of palaeovegetation, climatic reconstructions of the past, historical linguistics, as well as knowledge obtainable from other fields of science, have to be taken into account in forming theories of the spread of anatomically modern humans over all inhabited continents (reviewed recently, e.g., in Harpending et al. 1998; Renfrew and Boyle 2000; Underhill et al. 2001a; Cavalli-Sforza and Feldman 2003). Besides the effects of large demographic events, geographical clustering of the Y-chromosomal variation is influenced by random genetic drift, including founder effects and also by demographic (including social) behavior of men — the bearers of Y chromo-

some — and the society in general. It has been suggested that approximately 70% of modern societies practice patrilocality (e.g. in Seielstad et al. 1998). That means more men are considered to live closer to their birthplaces than women and local differentiation is enhanced, forming clinal distribution patterns of lineages in case of large and stable populations. This phenomenon has been used in interpreting the pattern of Y-chromosomal variation in Europe (Rosser et al. 2000) and in island Southeast Asia (Kayser et al. 2001). The examples of social aspects influencing distribution of Y lineages, in fact opposite to patrilocality, are sex-specific gene flow that accompanied the expansion of Europeans into Americas and Oceania in the past 500 years and strong introgression of European Y chromosomes with retention of indigenous mtDNA lineages that is seen in Polynesia (Hurles et al. 1998), Greenland (Bosch et al. 2003) and South America (Carvajal-Carmona et al. 2000; Carvalho-Silva et al. 2001).

2.6. Major branches of Y-chromosomal phylogenetic tree and the dispersal of modern humans

Evidence from Y chromosome shows that despite of uncertainty about the time to the most recent common ancestor (TMRCA) of Y chromosome, no ancient, more than 200 kya bifurcations of the Y chromosome lineages have been postulated anywhere in the world and the Y phylogeny roots in the Africa around 100 kya (Hammer et al. 1998; Underhill et al. 2001a; Underhill 2003). The two deepest branches — A and B (fig. 2) — both show a wide distribution in sub-Saharan Africa, though generally present at moderate or low frequencies (Underhill et al. 2001a; Semino et al. 2002). The reason, why the imprints of the Paleolithic events are faint in Africa is thought to be in a recent substantial expansion of hg E chromosomes, encompassing about 80% of the present-day African Y chromosomes, probably distributed by iron-working Bantu-speaking farmers from West Africa, starting about 3–4 kya (Underhill et al. 2001a).

Restriction of phylogenetically deepest lineages to Africa and evidence for an expansion out of Africa witnesses that modern diversity arose in Africa and replaced Y chromosome variants elsewhere in the World (Underhill et al. 2001a). Contemporary global Y-chromosomal variation is therefore quite reliably thought to descend from men (people), who migrated out of Africa about 60–50 kya.

Arhaeological evidence supports the theory that there were at least two distinct migrations out of Africa in 60–50 kya (Cavalli-Sforza et al. 1994; Lahr and Foley 1994). There might have been an early southern migration route (Lahr and Foley 1994; Stringer 2000; Walter et al. 2000), probably followed by a coastal route around the northern edge of the Indian Ocean before 50 kya and

a slightly later northern migration (Bar-Yosef et al. 1986; Clark and Lindly 1989) into Eurasia over Sinai, via the Levantine corridor.

In fact, there are several different hypotheses about the initial colonization routes, some of them giving more importance to the northern route (Underhill et al. 2001a) and others that signify the role of the southern route (Kivisild et al. 2003; Underhill 2003). There exists also a pincer model (Ding et al. 2000) of colonizing East Asia by two separate routes (northern and southern pincers), arguing that the present pattern of distribution of haplotypes (markers) suggests simple isolation by distance.

It has been stressed that the original founders diversified into lineages that display an irregular geographic distribution (Underhill 2003). For example, mutation M168 (fig. 2) represents a signature of the recent successful modern human migrations across Africa and beyond, as it is a central node at the root of all Out-of-Africa Y-chromosomal haplogroups, except of African-specific haplogroups A and B. Distribution (phylogeography) of lineages that descend from M168 permit to follow the major movements that have occurred after the humans left Africa. In other words, they represent the best approximation of the true coalescence tree, although an unknown number of branches may have been lost since then.

The majority of branches of the Y chromosome tree outside Africa are composed of a tripartite assemblage of the following haplogroups: a) C; b) D and E, and c) an overarching haplogroup F that defines the internal node of all remaining haplogroups from G to R (fig. 2).

Because the mutation defining haplogroup C (M130=RPS4Y) has not been observed in any African populations, this haplogroup has likely arisen somewhere in Asia after an early departure of modern humans from Africa, prior the arrival of them to Sahul in Southeast Asia. The most westernmost region where haplogroup C* has been detected is India (Kivisild et al. 2003). This lineage consists of several sub-lineages with irregular phylogeographic patterning, ranging from Central and North Asia to America and in the direction of Southeast Asia up to Australia and Oceania (Forster et al. 1998; Bergen et al. 1999; Karafet et al. 1999; 2001; Kayser et al. 2000a; 2001; Capelli et al. 2001; Underhill et al. 2001b; Lell et al. 2002).

Differently from hg C, haplogroups E and D share three phylogenetically equivalent markers indicative of shared heritage originating most probably in Africa. Some descendants with these mutations remained in Africa, giving rise to hg E (see fig. 3), which is most frequent and divergent in Africa, while others left it, to become a part of gene pool of early colonizers of Asia.

Subsequent geographic separation and accumulation of mutations gave rise to hg D (M174), which generally occurs at low frequencies throughout eastern Asia (fig. 3), except in peripheral locations like Tibet, Japan, and the Andaman Islands, where significant frequencies of D have been observed, probably because of founder effects (Su et al. 1999; 2000; Su and Jin 2000; Thangaraj et al. 2003; Tajima et al. 2004; Wen et al. 2004).

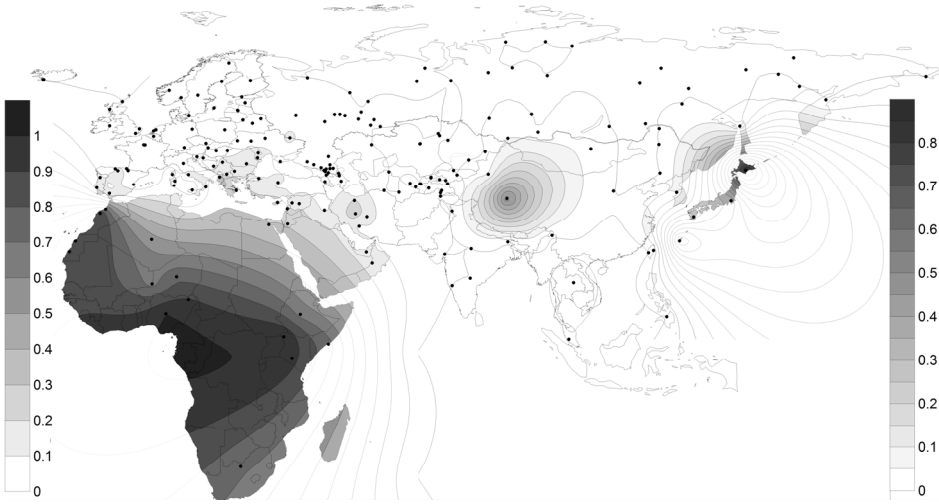

Figure 3. Spatial distribution of haplogroup E (in Africa and Western Eurasia) and D (in Eastern Eurasia) according to data of Helgason et al. (2000); Rosser et al. (2000); Semino et al. (2000a); Bosch et al. (2001); Wells et al. (2001); Zerjal et al. (2001); Laitinen et al. (2002); Passarino et al. (2002) Karafet et al. (2002); Al-Zahery et al. (2003); Barac et al. (2003); Francalacci et al. (2003); Maca-Meyer et al. (2003); Nasidze et al. (2003); Cruciani et al. (2004); Tajima et al. (2004). Frequency scale of hg E is shown on left and frequency scale of hg D is on right side of the figure.

The third major subclade of M168 lineages — superhaplogroup F — is characterized by mutation M89 at its root (fig. 2) and from it all other haplogroups deploy. F has been suggested to have evolved early in the diversification and migration of modern humans (Kivisild et al. 2003; Underhill 2003).

Later on, the ancestral trunk of F diversified into many branches by subsequent acquisition of mutations, giving rise to many region-specific haplogroups, such as J and G in Near and Middle East, I in Europe, H in Southern Asia, etc.

An expansion of F lineages (see fig.4) gave rise also to a population that acquired the M9 mutation (hg K), which defines another major bifurcation in the phylogeny (fig. 2). The branches of this clade probably migrated in different directions (North and East) and gave start to many separate and region-specific haplogoups in Eurasian continent and beyond.

Out of descendants of M9 lineage, hg L (M20) has greatest frequency in Southwest Asia and distinctive K lineages and M (M4, M5) haplogroup are restricted to Oceania and New Guinea (Kayser et al. 2000a; Hurles et al. 2003; Jobling and Tyler-Smith 2003), whereas hg O with its numerous sub-clades predominates in southern and southeastern Asia, reaching North China, Manchuria and some Siberian populations (Su et al. 1999; Su and Jin 2000; Tajima et al. 2002; Karafet et al. 2002). The population carrying M9 expanded also in direction of north towards Central Asia characterized by subsequent

mutations defining hg P, which encompasses distinctive eastward expanding hg Q (M242) characteristic to Siberian populations and Amerinds (Karafet et al. 1999; 2002; Hammer et al. 2001; Underhill et al. 2001a; Wells et al. 2001) and Eurasian hg R lineages that have expanded westward (fig. 4). Thus, one may speculate that multiple independent formations and fragmentations of populations carrying F-related lineages throughout most of Eurasia may have displaced the earlier hg C and D lineages towards the margin in many areas.

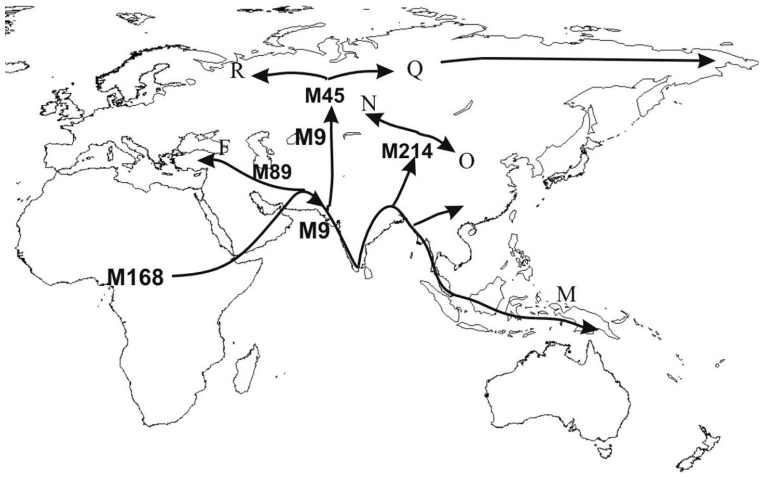

Figure 4. Schematic reconstruction of superhaplogroup F (defined by M89) origin, subsequent diversification of M9 lineages and their possible migration routes across the world. Adapted from Underhill (2003).

The previously described model presented by Underhill (2003) is similar to that suggested in Kivisild et al. (2003), according to which the initial coastal (southern route) migration(s) from Africa carried the ancestral Eurasian lineages first to the coast of the Indian subcontinent or that some/many of them actually originated there. Indians show the presence of diverse lineages of the three major Eurasian Y-chromosomal haplogroups C, F and K, although they have probably lost the fourth potential founder D, which is present in Andaman Islands population (Thangaraj et al. 2003). Next, the reduction of general package of four Y chromosomal (C, D, F and K) founders to two (F and K) occurred during the westward migration to western Asia and Europe.

After this initial settlement process, each continental region (including the Indian subcontinent) developed its region-specific branches. Western Asia and Europe have thereafter received an additional wave of genes from Africa, likely via the Levantine corridor, bringing hg E lineages, absent in India.

2.7. Distribution of the Y-chromosomal variation in Asia

Asia is a vast continent where genetic drift has played an important role in shaping the Y-chromosomal variation, affecting more seriously smaller populations. Therefore, the mean values of haplogroup distribution over the whole Asia, taken alone, are not informative for making conclusions about their phylogeographic spread: 15 of 18 major haplogroups are present in Asia, but their distribution is highly region-specific (Hammer and Zegura 2002). Here the attention is focused mainly to these Eurasian regions and Y-chromosomal haplogroups that have played role in colonizing Europe, while the more distant areas are considered only briefly.

2.7.1. Colonization of South and East Asia

There are two major models of migration routes of the initial peopling of East Asia. The first one argues that an early southeastern Asian spread via the southern route is a more likely scenario, followed by a northward migration. Some genetic surveys, such as the variation of autosomal microsatellite markers (Chu et al. 1998) and Y-chromosomal binary markers (Su et al. 1999; Jin and Su 2000; Su and Jin 2000) support this model. This result is concordant mainly with the distribution of major Asian hg O with larger diversity in southern *versus* northern regions and its northern sub-clades being a subset of the southern variation (Su et al. 1999). In contrast, the second model suggests a bi- and/or multidirectional route: one migration possibly through Central Asia and one through Southeast Asia (Ding et al. 2000; Karafet et al. 2001; Wells et al. 2001; Tajima et al. 2002; Jin et al. 2003). In previously named studies the main haplogroups for which the possible northern route is discussed are hg C with its very wide overall Asian distribution, together with hg D, the latter displaying a very specific and restricted distribution, absent or present in marginal frequencies in most populations of Asia, but frequent in Tibetan, Andamanese and Japanese populations.

2.7.2. Y-chromosomal landscape in Siberia and Central Asia

Large regions of North Asia are inhospitable and have never supported high population densities. Low population size leads to strong genetic drift and such regions show patterns of diversity that differ from that in more densely populated areas (Avise 2000). Characteristic features common to many native Siberian populations studied by Karafet et al. (1999; 2002); Derenko et al. (2002); Lell et al. (2002); Stepanov (2002) were pointed out by Karafet et al. (2002) as follows:

(i) only four major clades (N, C, Q and R) describe more than 96% of Siberian Y chromosomes.
(ii) many individual populations have a single predominant haplogroup (like 90% of Oroqens, 74% of Evens, about 70% of Eastern Evenks, 60% of Buryats and 52% of Mongolians belong to hg C, about 90% of Yakuts and 54% Eskimos to hg N3, 92% of Nganassans and 74% of Tundra Nenets belong to N2, 94% of Kets and 66% of Selkups belong to hg Q and 47% of Altaians to hg R1a).

Phylogeographic analysis of haplogroups with frequencies more than 10% in Native Siberians (Hammer and Zegura 2002) revealed that the two most frequent haplogroups were sub-clades of hg N: N3 with a frequency of 22.7%, was widely distributed within Siberia and northern Europe, whereas N2, a sister-clade of N3, with a frequency of 19.7%, had a much more spotty distribution which correlated with the spread of languages — 92% of the Siberians with this haplogroup are the Uralic-speakers. A common haplogroup present in Native Siberians is C3(xM48) defined by M217 (see fig. 2) present at frequency of 9.5% with one major sub-haplogroup C3c (13%; defined by M48) in mostly Altaic-speaking populations. Hg Q has quite restricted spread in Siberia, being highly frequent only in Kets and Selkups, while its sub-clade Q3 is predominant among Native Americans but absent in Asians. Hg R1a with a frequency of 10.3% was concentrated in the Altaian and some Northwest Siberian populations.

Y-chromosomal genetic diversity in Siberia is more structured according to language than geography (Karafet et al. 2002) — in contrast to that suggested for Europe (Rosser et al. 2000).

Similar pattern of marked distinction between populations with high or low settlement density can be seen also in Central Asia (Perez-Lezaun et al. 1999; Zerjal et al. 2002). According to Wells et al. (2001), more frequent haplogroups in Central Asian region are R (R1 and R1a), C3, J2 and, to a lesser extent, F*, E, K, O, L and R2.

Central Asia is located at the crossroad between West and East Eurasia. Therefore, Wells et al. (2001) have speculated in terms of putative importance of Central Asia as a starting point of different migrations to Europe, India, as well as to the Americas. Influence of the Near East Y chromosome pool to the western part of Central Asia was detected by the frequency gradient of haplogroup J, in particular as far as the present-day fertile Fergana Valley region populations are concerned, whereas the influence from East European steppes to eastern Central Asia can be deduced from the pattern of the distributon of hg R1a (Zerjal et al. 2002). In addition, Northeast Asian Y chromosome contribution to Central Asians can be inferred from the presence of hg C. Hence, Central Asia can be seen as a complex, multi-directional donor region of Y-chromosomal variation. Yet for some haplogroups, like for some derivates of M45 and for sub-haplogroup C3c, defined by M48, the extant Central Asian pool of paternal lineages can be seen as a source population for

expansion (Zerjal et al. 2002). Summing up, it appears that the Central Asian variation of the Y chromosome reflects its status as as a recent admixture zone of paternal lineages arriving from different directions/regions, rather than of the zone of their origin and initial spread. It can be stressed here that the latest mtDNA studies strongly support the former interpretation of the otherwise remarkable genetic diversity observed in Central Asian populations (Comas et al. 2004; Quintana-Murci et al. 2004).

2.7.3. Y-chromosomal variation in the Near East

Macro-haplogroup F has probably evolved in East Africa (see Chapter 2.6), from where it dispersed to western Eurasia around 50 000 years ago. This expansion is characterized morphologically and archaeologically by the first Upper Palaeolithic cultures, as well as evolutionary by the extinction of the western Asian and European Neanderthals (Underhill et al. 2001a; Klein 2003).

As it has been stressed above, the exact route of initial colonization of the Near East by anatomically modern humans is still controversial. Cavalli-Sforza and Feldman (2003) discuss weather the separation to southern and northern migration routes occurred already in Africa, or after the entry of the pioneer African colonists to West Asia and if there, then where did it happen.

The terms Near East and Middle East are used in population genetics literature in an overlapping meaning, often as synonymes. Therefore, to avoid confusion and to be more precise where it seems important, the Near East is defined here as the Levant and regions surrounding it, including Anatolia and Mesopotamia, and the Middle East as more eastern regions such as Iran and Afghanistan, being, however, well aware that the current literature is more relaxed in employing these two terms.

The Near Eastern region spanning from Zagros Mountains and northern Mesopotamia to Southeast Anatolia, called Fertile Crescent, is considered to be the place where agriculture first arose and from where it started to spread.The same region is often considered to be the starting-point for the Neolithic expansion about 10 000 years ago (outlined in Cavalli-Sforza et al. 1994). There are data, indicating that late Natufians were probably the earliest farmers in the Levant (Bar-Yosef 1998).

The Y chromosomal haplogroups most characteristic to the Near Eastern populations are J and E3b, which are also considered to be the most likely markers for Neolithic expansions (Cruciani et al. 2004; Luis et al. 2004; Semino et al. 2004). Hg J has most probably arisen in the Near East, where it has also the highest frequency and diversity, exhibiting a decreasing clinal pattern from the Near East to Mediterranean Europe, North Africa, Iran, Central Asia and India. Hg J has two larger sub-haplogroups J1 and J2 (fig. 2). Nebel et al. (2001) showed that two haplogroups, J1 and J2, constitute the major part of the

Near Eastern Y chromosome pool and suggested that J2 has originated in the northern part, and J1 in the southern part of the Fertile Cresent, from where they later differentially expanded. It was speculated that hg J1 has spread by two temporally distinct migratory episodes, the most recent one probably associated with the diffusion of Arab people (Nebel et al. 2001), while hg J2 distribution is consistent with Levantine/Anatolian dispersal route to southeastern Europe and may reflect the spread of Anatolian farmers (Nebel et al. 2001; Semino et al. 2004).

2.8. Y-chromosomal haplogroup variation in Europe

Y-chromosomal variation in European populations, mostly region or population-specific, has been studied by now quite in details. Thus, data on North (Lahermo et al. 1999; Helgason et al. 2000; Rootsi et al. 2000; Dupuy et al. 2001; Raitio et al. 2001; Zerjal et al. 2001; Laitinen et al. 2002; Passarino et al. 2002; Villems et al. 2002), North-West (Wilson et al. 2001; Weale et al. 2002; Capelli et al. 2003), Central and East Europe (Semino et al. 2000b; Passarino et al. 2001a; Stefan et al. 2001; Ploski et al. 2002; Kharkov et al. 2004; Kuzniar and Ploski 2004), South and West Europe (Semino et al. 1996; Belledi et al. 2000; Bosch et al. 2001; Passarino et al. 2001b; Scozzari et al. 2001; Francalacci et al. 2003; Maca-Meyer et al. 2003; Semino et al. 2004) have become available during a few recent years. Although the earlier studies exploited less markers, Europe is probably the most thoroughly studied wider area worldwide and, as a result, the phylogeographic pattern of Y-chromosomal variation in this continent is understood more precisely than that for other regions.

The most widespread early picture of European Y chromosomal landscape have been offered by two parallel surveys by Semino et al. (2000a) and Rosser et al. (2000), which both revealed similar clinal patterns for major European haplogroups. Among the two, Rosser et al. (2000) study has been carried out at somewhat lower phylogenetic resolution level, necessitating a need of updating phylogeographic coverage provided. The same, though at present in lesser extent, holds true also for the former study. For example, the spread of a major haplogroup 2 (nomenclature as used in Rosser et al. 2000) did not show any clinal pattern of variation in Europe. However, it was shown later on (YCC 2002) that this haplogroup is actually a paraphyletic group, whereas the constituents of it show much more differentiated patterns of spread. Indeed, in phylogenetically deeper or/and more recent studies (e.g. Semino et al. 2000a; Bosch et al. 2001; Cinnioglu et al. 2004; Semino et al. 2004; Cruciani 2002), several region-specific haplogroups, such as I (European specific), G (Near Eastern, Caucasian), B (sub-Saharan African) are dissected from the previous paraphyletic hg 2.

Semino et al. (2000a) found that more than 95% of studied European Y chromosomes can be grouped into 10 phylogenetically defined haplogroups.

Geographic distribution and age estimates were interpreted as testifying for two Paleolithic and one Neolithic migratory epizodes that have contributed to modern European gene pool. The majority of European Y chromosomes belong to hgs R1a, R1b, I and N3, which taken together, cover about 70–80% of the total Y chromosome pool (table 1). The remaining 20% of males belong to haplogroups J2, E3b or G.

According to the authors (Semino et al. 2000a), the distribution of M173 lineages (fig. 2) suggests that M173 is an ancient Eurasiatic marker that was brought by or arose in the group of *Homo sapiens sapiens* who entered Europe and diffused from east to west about 40 000 to 35 000 years ago, spreading the Aurignacian culture. This time estimate of initial colonization of Europe is concordant also with other genetic studies and archaeological data (Richards et al. 1996; 2002; Boyd and Silk 1997; Klein 2003). Recently, at least 35,000 years old jawbone of modern human fossil was found in Romania, in Transylvanian Alps, making it the earliest modern human remains found so far in Europe (Gibbons 2003).

Table 1 Frequencies (%) of major Y-chromosomal haplogroups in some European populations

Population	n	R1a	R1b [a]	I	N3	J2	E3b	G
Western Europe								
Catalan [1]	24	0	79.2	4.2	0	4.2	4.2	8.0
French [1]	23	0	52.2	17.4	0	13.0	8.7	0
Dutch [1]	27	3.7	70.4	22.2	0	0	3.7	0
Southern Europe								
Italians [1]	50	4.0	62.0	8.0	0	14.0	2.0	10.0
Croats [1]	58	29.3	10.3	44.8	0	5.2	6.9	1.7
Albanians [1]	51	9.8	17.6	19.6	0	23.5	21.6	2.0
Eastern Europe								
Polish [1]	55	56.4	16.4	23.6	0	0	3.6	0
Ukrainian [1]	50	54.0	2.0	18.0	6.0	6.0	4.0	4.0
Northern Europe								
Norwegian [2]	72	23.6	27.8	40.3	6.9	0	0	0
Danes [3]	194	16.5	36.1	38.7	0.5	2.6	2.1	0
Finns [4]	39	7.9	0	28.9 [b]	63.2	0	0	0

Data from: [1] Semino et al. (2000a); [2] Passarino et al. (2002); [3] Sanchez et al. (2003) [4] Zerjal et al. (2001)

[a] frequencies of R1b are deduced, defined here by M173(xSRY-1532), as studied western Europeans from this clade have been shown to share the additional mutation M269 (Cruciani et al. 2002);

[b] frequency of haplogroup I is deduced and defined by the characteristic STR pattern of haplogroup 2 Y chromosomes.

About 50% of European Y chromosomes share the M173 marker that defines R1 clade and consists of two separate branches harboring contrasting geographic distributions. One, R1*, is defined by M173(xSRY1532), here deduced to R1b according to Cruciani et al. (2002), where all studied western Europeans from this clade have been shown to share the additional mutation M269 (fig. 2). Hg R1b shows decreasing frequency from west to east, while the second — R1a — is defined by SRY 1532 plus M17 (R1a1) and is showing opposite frequency gradient in Europe, with its maximum frequency in eastern Europe, particularly in Slavic populations (table 1).

Semino et al (2000a) attributed the spread of R1a to the post-LGM recolonization of Europe from the refugial area in the territory of the Ukraine. An alternative possibility, linking the spread of R1a to the movement of the Kurgan people from north of the Caspian Sea in a much more recent time scale, has been suggested by Rosser et al. (2000).

The distribution of hg R1b covers actually an area wider than Europe, but the centre of its cline in Europe lies in western Europe, revealing post-LGM spread from the Iberian refuge area as was speculated by Semino et al. (2000a) and is associated with specific 49a,f TaqI haplotype 15 (Semino et al. 1996), while in the Balkans, the Caucasus and in Anatolia, R1b individuals mostly possess 49a,f TaqI haplotype 35 (Semino et al. 1996; Cinnioglu et al. 2004).

The polymorphism M170 (hg I, see fig. 2) represents another putative Palaeolithic mutation which age has been estimated to be about 22 000 years and it has been proposed by Semino et al. (2000a) that M170 originated in Europe in descendants of men who arrived from the Near East about 25 000 years ago, associated with the arrival of the Gravettian culture.

During the LGM, many regions of Europe, in particular the northern areas became unsuitable for human occupation (Peyron et al. 1998; Kageyama et al. 2001) and were largely uninhabited for many thousands of years. After climatic improvement, repopulation has started, most likely as expansions from isolated population nuclei from different refugial zones in Europe — Iberia, the present Ukraine and, perhaps, from the northern Balkans (Dolukhanov 2000).

Haplogroup N3 defined by *Tat C* allele (Zerjal et al. 1997) is present in northern and eastern, but missing in western and southern Europe (table 1, see also chapter 2.7.2), being frequent also in northern Asia (Zerjal et al. 1997; Rootsi et al. 2000; Rosser et al. 2000; Zerjal et al. 2001; Semino et al. 2000a; Laitinen et al. 2002; Villems et al. 2002). It is quite remarkable that none of the mtDNA haplogroups has even a remotely similar phylogeography.

Besides previously discussed lineages**,** there are several haplogroups marked by M35, M172, M89, and M201 (respectively E3b, J2, F and G) with clines of frequencies decreasing from the Near East to Europe (Semino et al. 2000a; Cruciani et al. 2004; Semino et al. 2004). Therefore, haplogroups E3b, J2 and G (table 1) have been considered to represent the male contribution of a demic diffusion of farmers from the Near East to Europe, accounting for 22% of the present-day European gene pool (Semino et al. 2000a). Furthermore, their

observed frequency patterns revealed that the putative contribution of the Neolithic farmers to the European gene pool is more pronounced along the Mediterranean coast than in continental Europe (Semino et al. 2000a; 2004; table 1).

It has been speculated by Semino et al. (2004) that the distribution of J-M172(xM12) haplogroup is consistent with its spread to Europe through the Levantine corridor, congruent with the distribution of mitochondrial haplogroups J, K, T1 and HV (Richards et al. 2000; 2002) and that haplogroups E-M78 and J-M12 trace the diffusion of people from the southern Balkans to the west (Semino et al. 2004).

It is generally accepted that agriculture arose in the Near East. There exist two contradicting models about the mechanism of dispersal of farming. Population geneticists often debate fiercely either in favour of one or another model, while their experimental data may not differ that much at all. The models are typically presented as follows:

1) The demic diffusion model proposed by Ammerman and Cavalli-Sforza (1984) postulates that extensive migrations of Near Eastern farmers brought agricultural techniques to Europe. Extreme variants of this model tend to suggest that there was little admixture between the expanding Neolithic farmers with the Mesolithic Europeans and that the latter were largely replaced, so that a large proportion of the present-day European gene pool derives from Neolithic migrants. This model has been supported by Piazza et al. (1995), Cavalli-Sforza and Minch (1997), Chikhi et al. (1998; 2002) and Barbujani et al (1998).

2) The cultural-diffusion model by Dennell (1983) and Zvelebil (1986; 2000)**,** in which the transfer to food production occurred without significant population movements and the majority of the genetic diversity within Europe should have its roots in the Paleolithic Europeans.

Black-or-white extremes of these two models should be considered as over-simplistic. It must be stressed here that even in its "mature classical form" (Cavalli-Sforza et al. 1994), the demic diffusion model of the spread of agriculture predicts that the "Neolithic farmers" account for about 27% of the present-day gene pool of Europeans only — a result that does not lie far apart from most of the recent mtDNA and Y-chromosome-based estimates.

As far as the Y chromosomal evidence hints so far, the demic diffusion may have played more significant role in southern Europe in agreement with its geographic proximity to the Near East. For example, frequency cline of hg J2 from the Near East can be interpreted in terms of the arrival of agriculturalists (Semino et al. 2000a; King and Underhill 2002; Semino et al. 2004). Yet simple frequency clines cannot distinguish between events that had taken place, e.g., in post-glacial time or in Holocene and should be, in ideal, supported by reliable time estimates since the expansion, the latter derived largely from diversity estimates (Semino et al. 2004). For example, spread of hg J sub-clades may have occurred by a variety of mechanisms that involved gene flow(s) but different from demic diffusion — such as leapfrog spread, maritime spotty

colonization alongside the European Mediterranean coasts etc. (Richards 2003; Di Giacomo et al. 2004).

In distant parts of Europe cultural transmission was probably more important, in particular in the northern regions where agriculture was developed much more recently and haplogroups that are considered to be markers of Neolithic movements are practically missing (table 1).

As already mentioned above, the range of "Neolithic genes" in the present-day European gene pool has been estimated to be about 20–30% in most studies, both for Y-chromosomal and mtDNA lineages (Richards et al. 2000; Rosser et al. 2000; Semino et al. 2000a; Bosch et al. 2001; Richards 2003). However, there are also radically different estimates in literature. In case of Chikhi et al. (2002), at least twice higher (50–65% at least) Neolithic contribution was suggested by an incorrect presumption that Basques and Sardinians, with their fashioned by founder effect(s) and random genetic drift patterns of frequencies of genetic markers, represent "true" Palaeolithic Europeans and that only lineages present in these isolates, can be used to measure "Palaeolithicity" of the European gene pool.

3. AIMS OF THE STUDY

As it was already explained above, uniparentally inherited mtDNA variation in humans has been extensively studied for more than a decade, while exploring the other uniparentally transmitted genetic system — Y-chromosomal haplogroup variation — started later and significant progress has been made largely during the last five years, when many new and informative biallelic markers have been discovered, described and screened in large-scale population studies. The obtained results cover an increasing number of populations in different geographic areas and offer new possibilities for a deeper insight into the details of spread patterns of individual lineages.

The first goal of the present study was to improve the knowledge about Y-chromosomal haplogroup variation in Europe and its surrounding regions, because sound phylogeographic reconstructions are feasible only with a solid empirical background at hand. In Europe, quite extensive data exists for many southern European populations, while Balkan region, northern and eastern Europe has been so far studied less thoroughly. Our purpose was to fill the gaps in datasets (in particular as far as the easternmost Europeans — Volga-Uralic region populations were concerned) and use the acquired data for deeper phylogeographic studies.

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 and Y-chromosomal lineages are spread among the Saami in a wider Eurasian context — where did these lineages possibly arise, how did they reach the northernost Fennoscandia and are the Saami indeed "outliers" among European populations, or simply a small distinct part of the European unity.

Although there is a sound evidence that the majority of the present-day European genes descend from indigenous Palaeolithic ancestors, it is to be expected that the pre-LGM landscape of the spread of genetic variation has been profoundly re-shaped during and after the LGM (Richards et al. 2000; Semino et al. 2000). Our aim was to apply the phylogeographic approach on Y-chromosomal haplogroup I, the only known Y-chromosomal haplogroup that has most probably arisen in Europe in Palaeolithic times and is still common and widespread there.

Distribution of haplogroup I was intriguing with its two high frequency peaks in distant parts of Europe (the Balkan region and Scandinavia) and our study concentrated on achieving a better phylogenetic and phylogeographic resolution of this haplogroup, informative for the reconstruction of long-distance gene flows in space and time.

4. SUBJECTS AND METHODS

4.1. Subjects

The experimental basis of current thesis employs the analysis of Y-chromosomal variation of different population samples: 303 Portuguese, 132 Maderians, 121 Azoreans, 201 Cape Verde Islanders, 55 Slovenians, 457 Croats, 100 Bosnians, 51 Albanians, 361 Romanians, 377 Estonians, 86 Latvians, 93 Polish, 53 Czechs, 70 Slovaks, 113 Hungarians, 535 Ukrainians, 144 Swiss, 179 French, 225 Swedes, 35 Saami, 60 Moldavians, 79 Gagauz, 147 Byelorussians, 766 Russians, 89 Udmurts, 83 Mordvin, 110 Komis, 80 Chuvashes, 126 Tatars, 61 Nogays, 138 Adygeis, 70 Karachais, 89 Armenians, 64 Georgians, 47 Ossetians, 122 Yakuts, 93 Turks and 83 Iranians.

Blood samples were obtained from healthy, unrelated volunteers with informed consent. DNA was extracted using the phenol-chloroform method, as used by Sambrook (1989). Two sets of Estonian samples (167 and 210) and Estonian Russians (97) were collected by members of Laboratory of Evolutionary Biology of Tartu University, other samples from different populations were gained in collaboration with Institute of Biochemistry and Genetics in Ufa, Research Center for Medical Genetics in Moscow, Institute for Anthropological Research in Zagreb, Institute of Anthropology in Zurich, Institute of Cellular Biology and Pathology "Nicolae Simionescu" in Bucharest, Human Genetics Laboratory, Center of Macaronesian Studies, University of Madeira, Laboratoire d´Etude du Polymorphisme de l´ADN, Faculté de Médecine in Nantes, Department of Molecular Genetics of Comenius University in Bratislava, Institute of Forensic Sciences in Budapest, Gotland University in Visby, Department of Biology, Faculty of Natural Sciences of Tirana University and Medical Academy of Latvia in Riga.

4.2. DNA typing

DNA samples were amplified by PCR reaction using primers for specific markers. For most of markers further genotyping was performed either by RFLP analysis or sequencing the polymorphic sites of markers. Only in case of YAP insertion and 12f2 deletion genotyping was done directly by detecting the difference of product size in 2% agarose gel. Markers were typed in hierarhical order and haplogroup affiliations were established by combination of derived states of typed markers. In the earlier study (IV) samples were typed (markers M9, M89, 12f2, YAP, Tat, M20, 92R7 and SRY-1532) and haplogroups were named according to the nomenclature of Jobling et al. (1997), but in more recent works (I-III) the nomenclature of YCC (2002) was used. Typed markers were (ref. II, III) M9, Tat, SRY-1532, 92R7, M89, P43, YAP, 12f2, M52,

M130, M170, M173, M178, M201, M269, M242. In case when a population was investigated in earlier study for lesser markers, additional markers were typed (ref I–III) to achieve the resolution of present nomenclature. In ref. 1 marker M170 was studied and samples showing the derived C-allele were further typed for markers M253, P37, M223, M227 and M26 defining the subclades of haplogroup I.

STRs were typed (ref. I, II, III, IV) by using the Automatic Laser Fluorecence Express DNA sequencer and detection of repeat numbers for individual markers was performed by Allele Links version 1.00 software (Pharmacia Biothec). Five (ref. IV) or six STRs (ref. I, II) were typed (DYS19, 388, 390, 391, 392,393) for majority of studied samples, but for I1b haplogroup samples (ref. I) an additional marker was investigated (YCAIIa,b). Eight STR loci (DYS19, 388, 389I, 389II, 390, 391, 392, 393) were typed and analyzed in ref. III.

Detailed descriptions of methods are given in references I–IV.

In ref. I populations — Portuguese, Swiss, Maderians, Azoreans, Cape Verde Islanders, Slovenians, Croats, Bosnians, Albanians, Romanians, Estonians, Latvians, Polish, Czechs, Slovaks, Hungarians, Ukrainians and Turkish, French, Swedes, Saami, Moldavians, Gagauz, Byelorussians, Russians, Udmurts, Mordvin, Komis, Chuvashes, Tatars, Nogays, Adygeis and Karachais were typed by the author or in collaboration with H. Help, M. Bermisheva, O. Balanovsky and A. Pshenichnov. Other samples used in this study were typed in Pavia or Stanford University. The author performed the further analysis of the data. Age estimates were calculated in collaboration with L. Zhivotovsky.

For ref. II 16 bi-allelic markers were typed by the author or in cooperation with H. Help and M. Bermisheva. Analysed populations were: Saami, Swedes, Estonians, Latvians, Poles, Maris, Mordvin, Komis, Udmurts, Chuvashes, Tatars, French, Hungarians and Yakuts.

In ref. III Ychromosome variation in 457 Croatian samples was studied by L. Barac and M. Pericic with assistance and help of the author of present thesis. The author participated also in later analysis of data.

In ref. IV Estonians, Hungarians, Russians, western Slavs (Poles, Czechs, Slovaks), Trans-Caucasians (Armenians, Georgians, Ossetians) and Volga-Ural populations (Udmurts, Chuvashes, Tatars) were studied by author for biallelic markers and samples with Tat C-allele were typed for 5 STRs to estimate diversity level in haplogroup 16 for populations of different regions.

4.3. Data analysis

To visualize the distribution and frequencies of haplogroups the maps were obtained (ref. I), applying the frequency data in Surfer (version 7) software (Golden Software, Inc.) or frequency pie charts shown on contour maps (ref. IV).

Haplogroup (based on bi-allelic markers variation) and haplotype (based on STR allele variability) frequencies and diversities (calculated as in Nei 1987) of particular haplogroups and their sub-clades (ref. I, IV) were calculated and compared, including into the study also data published earlier by others.

Populations (more precisely: their Y-chromosomal variation) were compared by the use of Principal Component (PC) analysis, based on haplogroup frequencies (ref. II).

The phylogenetic networks relating different haplotypes (as described in Bandelt et al. 1995 and 1999) were constructed. The networks are constructed on basis of STRs allele variants (repeat number of STR repeat units) or combined with data from bi-allelic markers (ref. I, III). For smaller datasets (less than 100 samples) the reduced median network (Bandelt et al. 1995) can be used but for larger datasets of several hundred samples, the median joining algorithm (Bandelt et al. 1999) was applied (ref. I, III).

The time estimates and divergence times of haplogoup I and its sub-clades (ref. I) were calculated by the use of method of Zhivotovsky et al. (2001, 2004) for variability of 6 STRs (DYS19, 388, 390, 391, 392, 393).

5. RESULTS AND DISCUSSION

In the List of Publicatins the articles are ordered according to their publishing time. In Results and Discussion section however, the referred articles are grouped according to the surveyed themes and subjects.

Our knowledge of the present-day phylogeography of Y-chromosomal lineages is rapidly increasing and many regions are already quite well covered. Nevertheless, data is still inadequate for others and it is quite often so that even if data does exist, the sample sizes involved are too small to make firm conclusions, in particular in seeking for the presence of less frequent haplogroups in a particular area. Furthermore, because of rapid progress in Y-chromosome research, several larger data sets published earlier, are by now only of limited value because of their less deep phylogenetic resolution compared to the present-day standards. Therefore, the initial purpose of our studies was to investigate so far poorly described regions/populations, to obtain data of sufficient phylogenetic resolution and to apply the phylogeograpic approach in order to study the spread of particular Y-chromosomal lineages in Europe and beyond.

5.1. Y-chromosomal variation in Croatians (ref. III)

In the analysis of European and circum-European populations by 11 SNP markers, only haplogroups defined by alleles of derived status showed distinctive phylogeographic patterns, while complex “macro-clades”, such as former haplogroups 2 (clade F in YCC 2002) and 26 (clade K in YCC 2002), showed a relatively uniform distribution over the continent (Rosser et al. 2000). Recently, new informative downstream markers allowed to improve the resolution considerably.

Making use of this progress in phylogenetic resolution, we have studied Croatian Y chromosomes and found that nearly half of Croatian samples (comparable to Yugoslavian samples in Rosser et al. 2000) which have been so far assigned to a poorly characterized superhaplogroup 2, belong to distinct haplogroups I or G and only 1.8% of individuals remain to super-haplogroup F (ref. III).

Altogether, 16 biallelic markers and eight STRs were used to analyze the Y-chromosomal diversity in the Croatian mainland (n=109) and four Adriatic island populations (Krk, Brač, Hvar and Korčula; n=348). The frequencies of nine haplogroups in five different subpopulations were established (table 1, ref. III). It turned out, that among Croatians, the prevailing haplogroup is I (M170), both in the Croatian mainland (37.6%) and in the islands (reaching 65.9% in Hvar). The second most frequent haplogroup in mainland and island populations is R1a, showing frequency distribution opposite to haplogroup I, being thus more frequent in the north of the country. The spread of R1a haplogroup in

East Adriatic and among the Slavonic-speaking Balkan populations in general can be, at least partially, related to the migration of Slavs to the region about 1400 years ago, because this haplogroup has characteristically high — about 50% — frequency in East Slavs (Semino et al. 2000a; Roser et al. 2000).

Meanwhile, the observed overall frequency of haplogroups E, G and J (12.5%) in the Croatian sample is low — a result that suggests only a minor genetic impact from the Near East.

In one of southern Dalmacian islands (Hvar), halogroup P*(xM173) in ref. III (samples later typed as hg Q, our unpublished data), a very unusual for Europeans variety of the Y chromosome, was found at a surprisingly high frequency (14%), suggesting a connection with Central Asian populations. This finding is particularly interesting, bearing in mind that it was found earlier that the maternal pool of Hvar is rich in mtDNA haplogroup F that is virtually absent in Europeans but typical and frequent in Central Asians as well as elsewhere in eastern Asia (Tolk et al. 2001).

5.2. The Saami: their position as so-called genetic "outliers" among European populations (ref. II)

Because of the persistent ambiguity in the origin of the Saami population, our laboratory, together with many colleagues outside, investigated their haploid genomes in a comprehensive Eurasian context (ref. II). As the mtDNA heritage and variation 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 Y-chromosomal haplogroups among the Saami.

5.2.1. MtDNA variation of the Saami

The analysis of the Swedish Saami samples and reanalysis of previously published mtDNA HVR I sequences of several other Saami populations from Finland and Norway confirmed that the "outlying" status of the Saami is caused by relative proportions of haplogroups in their mtDNA pool and not by distinctive, profoundly different maternal lineages compared to other Europeans. It has been also shown that there is little, if any, historic gene flow from Samoyedic- and Ugric-speaking populations from Siberia to the maternal gene pool of the Saami. Terefore, the distribution of the Saami mtDNA haplogroups is best seen as a restricted subset of the European mtDNA pool. It consists predominantly of two haplogroups that are widely spread in Europe, V and U5 (table 1 and 2, ref. II), which cover together more than 80% of the Saami mtDNA lineages. The rest of the variation is shared by European-specific haplogroups H, W and T. Only a small fraction of variants (about 5%) indicates an

admixture of the Saami mtDNA pool with eastern Eurasian lineages D5 and Z1. The most plausible explanation for the exceptional frequency pattern of mtDNA haplogroups among the Saami is that genetic drift, including bottleneck(s) and subsequent founder-effect(s), have had a decisive role in shaping the mtDNA variation of the Saami population. This phenomenon is particularly intriguing when we compare the mtDNA data with the Y-chromosomal pool of the Saami.

5.2.2. Y-chromosomal variation of the Saami

In contrast to their maternal gene pool, haplogroup composition of paternal lineages of the Saami population does not differ profoundly from that characteristic to their Finnic-speaking neighbours (fig. 2B and table 3, ref. II). Similarly to other northeastern Europeans, three major haplogroups — N3 (haplogroup 16 or Tat C allele in earlier literature, including in ref. IV), I and R1a — comprise about 80% of the Saami Y-chromosomal gene pool. These three haplogroups are nearly equally distributed among different Saami populations, which strongly supports their common paternal descent. Haplogroups J and E were found only among the Kola Peninsula Saami, suggesting their admixture with neighboring Russian-speaking population. This admixture is likely relatively recent, possibly linked to extensive mining industry, military and harbor buildup during the last centuries. Compared to Scandinavian Germanic-speaking populations — Swedes and Norwegians — where haplogroup N3 is present at frequencies less than 10% (table 3, ref. II), the Saami paternal pool is rich in it (37–55%). This feature brings Saami paternal lineages unambiguously together with their geographically close linguistic relatives (but see also Chapter 5.4. below).

The second most frequent haplogroup among Saami is I, encompassing about one third of their Y-chromosomal lineages. Contrary to haplogroup N3, it is present at high frequencies also among Germanic-speaking populations of Fennoscandia — Norwegians, Swedes, and is well present also in Finns. It suggests that haplogroup I may represent the heritage from the first settlers of Fennoscandia and that people carrying this haplogroup probably arrived to Fennoscandia from the west (fig. 4B ref. II).

Other more frequent haplogroups among Saami are R1a (11%) and R1b (4%). The former is frequent in eastern Europe and the latter is the most typical Y-chromosomal variant in western European populations (table 3, ref II).

Although Saami share haplogroup N3 with many Siberian populations, the fact that other haplogroups: N2, Q and C that are widely present in different Siberian populations (table 3, ref II) are totally absent in the Saami paternal gene pool suggests, that recent gene flow from Siberian populations to the Saami is unlikely. In this context, conclusions drawn from maternal and paternal heritages of the Saami, are congruent.

5.3. Phylogeography of haplogroup I sub-clades — an example of postglacial re-colonization of Europe (ref. I)

The knowledge gained from the previous studies (ref. II and III) was a starting-point to focus on Y-chromosomal haplogroup I, the only major clade of the Y phylogeny that is widespread over Europe but virtually absent elsewhere (ref. I).

It has been shown earlier that high frequency of hg I is characteristic for two distant and distinct regions — around the Dinaric Alpes (ref. III; Semino et al. 2000a) and in Nordic populations of Scandinavia (ref. II; Semino et al. 2000a; Passarino et al. 2002). The overall distribution of hg I in Europe is shown in figure 1 panel B (ref. I).

In this study 7574 individuals from Europe and surrounding regions were assessed for the marker M170, which defines hg I.

1104 Y chromosomes from 48 European and 12 populations from surrounding regions which showed the derived M170 C — allele, were further genotyped with a set of markers (M253, P37, M26 and M223). These markers define distinct sub-clades of I, respectively I1a, I1b, I1b2 and I1c. Frequencies of sub-clades are reported in table 1 (ref. I).

Phylogeny of studied markers in wider context of superhaplogroup F according to nomenclature of YCC (2002) is presented in figure 1 panel A (ref. I). Thanks to new informative markers the improved resolution of phylogeny of hg I enables to reveal distinct phylogeographical patterns of sub-clades I1a, I1b and I1c, which jointly cover about 95% of hg I individuals.

Sub-clade I1a is widely spread in northern Europe with its highest frequencies in Scandinavia: in Norwegians, Swedes and Saami, accounting for 88–100% of hg I individuals in these populations and showing rapidly decreasing frequency towards both the East European Plain and the northwestern coastal areas of Europe (fig. 1 panel C, ref. I). Combined analysis of STR diversity and relative portion of I1a sub-clade among all I lineages suggests that France or possibly more precisely — the Francocantabrian refugial area — could have been the source region of the spread of I1a during the post-LGM re-colonization of Europe. The same may apply to the spread of the less common sub-clade I1c. This scenario is also supported by high positive correlation (0.75) between the geographic distributions of I1a and I1c. I1c covers a wide range in Europe, with the highest frequencies in Northwest coastal Europe and lower frequency elsewhere (fig.1 panel D, ref. I).

A totally different distribution pattern can be seen for I1b*, which is the most frequent haplogroup I clade in eastern Europe and in the Balkans. It reaches its highest frequencies in Croatian and Bosnian populations, encompassing almost 80–90% of hg I there (fig. 1 panel E, ref I). When comparing frequencies in different regions of Croatia (ref. III), clear and significant difference between the three southern islands with higher frequency and mainland and northern island Krk with lower frequency, became apparent. More than

a half of Croatian hg I individuals — 126 out of 221 (57%) — share an identical STR haplotype, which was named as the Dinaric Modal Haplotype (fig. 2B, ref. III). This haplotype was not present in 102 hg 2 chromosomes (according to nomenclature of Jobling) reported by Helgason et al. (2000), the most frequent among them was labeled as the Nordic Haplotype (ref. III).

Phylogenetic network of hg I STR haplotypes (fig.2 in ref. I, fig. 5 in the thesis) points to characteristic haplotype patterns in different sub-clades that allows identify possible founder haplotypes for the sub-clades.

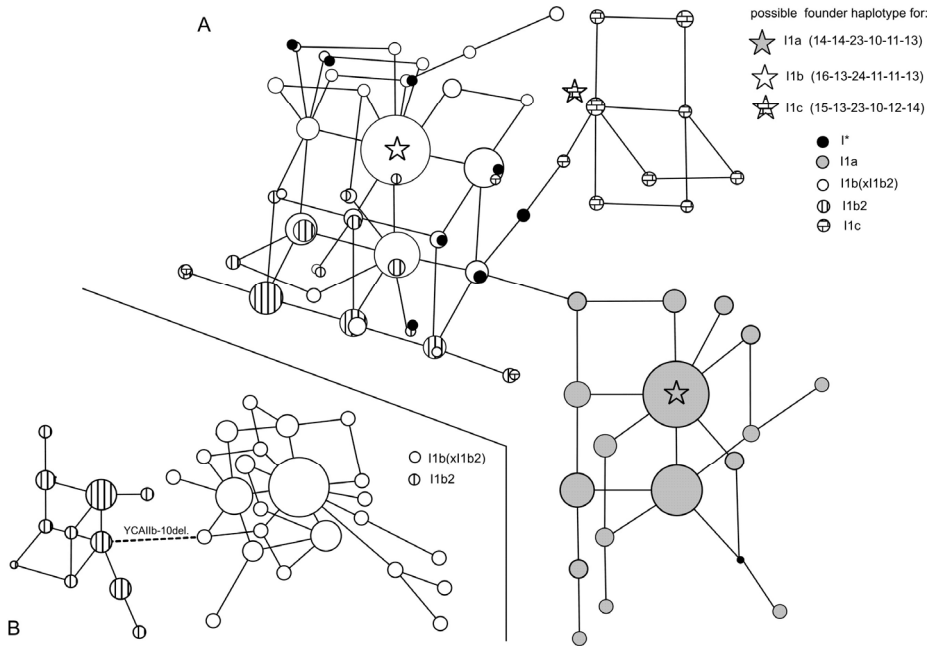

Figure 5. Phylogenetic network of haplogroup I.
A: Median joining network of combined haplotypes of six STR loci (DYS19, 388, 390, 391, 392, 393) in 25 populations/ 533 individuals (Norwegians, Estonians, Saami, Swedes, Hungarians, Czechs and Slovaks, Poles, Ukrainians, Croats, Bosnians, Macedonians, Albanians, Greeks, Moldavians, Gagauz, Turks, Italians, Sardinians, French, Dutch, Andalusians, Bearnais, Basques and Swiss). Only haplotypes with frequency >1 were used. Nodes indicate haplotypes with sizes proportional to their frequency (smallest node corresponds to 1 individual — only in case of overlap between subclades, otherwise haplotypes with frequency >1 are presented). Haplotypes of different sub-clades are indicated with different color patterns.

B. Median joining network of I1b lineges based on seven STRs (DYS 19, 388, 390, 391, 392,393 and YCAII a,b). I1b(xI1b2) and I1b2 lineages are clearly differentiated by the deletion of 10 repeats at YCAII b locus in M26 lineages.

The most frequent haplotype of I1a sub-clade (14-14-23-10-11-13) in ref. I corresponds to the earlier named Nordic Haplotype and the dominant in I1b* haplotype is the so-called Dinaric Modal Haplotype (16-13-24-11-11-13), described in ref. III. The possible founder lineage for the third sub-clade, I1c (15-13-23-10-12-14), was revealed (ref. I).

I1b* (YCAII a,b mostly 21,21 alleles) and I1b2 (YCAII a,b; 21,10 alleles) lineages are clearly separated from each other by the observed difference in YCAII a,b haplotypes (fig. 5B in the thesis; fig. 2 in ref. I). These sub-clades differ by 10 repeats in YCAII b allele length that is probably a result of one single mutational event rather than step-by step deletions, in particular because no intermediate repeat variants have been detected.

The estimates for possible expansion times suggest that the expansion phase of I1a and I1b occurred around the early Holocene and only the less frequent sub-clade I1c (table 3, ref. I) shows an earlier age for its STR variation, suggesting that the corresponding mutation has arisen earlier.

High frequency combined with high diversity of sub-clade I1b in Croatian population (both mainland and island populations) hints that during the LGM, there might have been a refugium close-by. According to our knowledge, placing of the western Balkans on the list of the human refugia during the LGM has not been confirmed so far unambigously by archaeologists. However, the northern part of the Adriatic Sea, including the Dalmatian Islands, was at that time a part of dry land, being covered by water only much later, at the boundary of Holocene (fig. 1 in ref. III). Therefore, one may speculate that the wealth of the traces of human occupancy of the area lies at present submerged. Nevertheless, it is justified to suggest that the present-day western Adriatic was the reservoir of M170 (I1b) lineages, as well as a starting point for spread of these lineages during the postglacial re-colonization of Europe. Meanwhile, the star-like pattern of both, I1a and I1b* STR haplotypes, might be explained by simultaneous re-colonization of Europe from different refugia.

I1b* sub-clade dissipates very rapidly west of the Balkans, being virtually absent among Italian, French and Swiss populations, but extends eastwards at notable frequencies, mostly in the north Balkans and among Slavic-speaking populations, including more eastern Ukrainians. This finding suggests that I1b* may have expanded from a glacial refuge area, what may have located in the Balkans. As indicated above, there is only limited archaeological evidence for such refugium in this region at present. Nevertheless, data on re-occupation of northern Europe from the Balkan region by other mammals like brown bear *Ursus arctos* (Taberlet and Bouvet 1994) and European hedgehog *Erinaceus europeus* (Hewitt 2000), birds — European great tit *Parus major* (Kvist et al. 1999; Kvist 2000) and insects — meadow grasshopper *Chortippus parallelus* (Hewitt 2000), supports its existence indirectly.

It seems somewhat less likely though not excluded that I1b* was preserved, during the LGM, in an area of much better documented Periglacial refugium in the present-day Ukraine. It appears less likely because (a) not only its

frequency, but also diversity is higher in the Adriatic region (table 2 in ref I); (b) a branch of I1b* — I1b2-M26 — has a clearly western pattern of spread, being totally absent in Ukrainians (fig. 1F, ref. I).

On the other hand, a clearly visible difference that can be observed in distribution patterns of I1b* and an outshoot of it — I1b2 — suggests that their separation may have occurred even before the LGM, whereas isolation, genetic drift during the LGM, re-colonization and an unknown number of putative more recent demographic events created the pattern that one observes among the extant populations.

Meanwhile, the extremely high incidence of I1b2 among Sardinians (about 40%) can be explained by the presence of carriers of I1b2 lineage among the first inhabitants of the island early in Holocene and by the influence of genetic drift afterwards.

Certain extent of similarity in distribution patterns of some mtDNA haplogroups, in particular V (Torroni et al. 1998; Torroni et al. 2001), with Y-chromosomal hg I sub-branches, has been suggested (ref. I).

5.4. The Y-chromosomal landscape in northern and eastern Europe (ref. II and IV)

The predominant haplogroup among the Saami, N3, has a nearly uniform circum-arctic distribution all over northern Eurasia (table 3 in ref. II and fig. 6 in ref. IV, fig. 6 in the thesis). Its spread is not restricted to any linguistically defined set of populations: it can be found frequently among many Uralic, Indo-European (East Slavonic and, in particular, Baltic), Chukchi-Kamchatka and Altaic-language speaking populations of the region. Hg N3 is particularly frequent among many populations of North Siberia, encompassing nearly 90% of the paternal gene pool of Yakuts and 40% in Nenets.

In Europe, this haplogroup is well present among eastern European populations of the Volga Basin (both in Finno-Ugric Komis, Maris, Mordvin, Udmurts and Turkic-speaking Tatars and Chuvashis) and in western Finno-Ugric Finns, Saami, Karelians and Estonians.

In this context, it is of a particular interest that haplogroup N3 is represented with equally high frequency (30-40%) among Latvians and Lithuanians (ref. II; Zerjal et al. 2001; Laitinen et al. 2002), who linguistically comprise the Baltic language branch of the Indo-European languages. Note that in the Indo-European language tree, the Baltic group (Latvians, Lithuanians and the long extinct Prussians) form a sister group to the Slavonic group of languages.

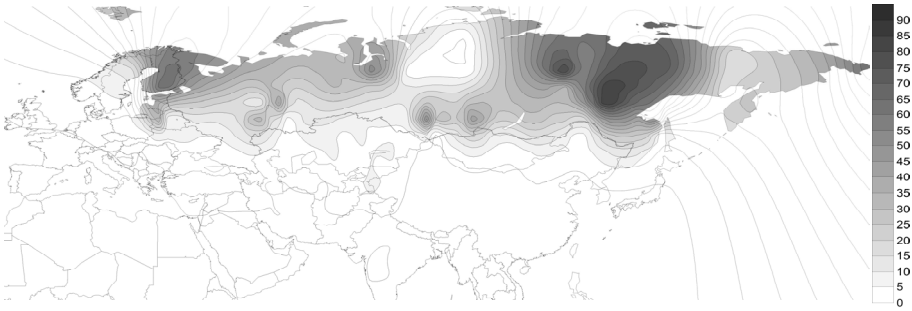

Figure 6. Circum-artic distribution of haplogroup N3 in Eurasia. Map is based on data from ref. II, IV and from Zerjal et al. (1997, 2001); Rosser et al. (2000); Wells et al. (2001); Karafet et al. (2002).

The north-south variation of hg N3 frequency from the Arctic Sea to Lithuania is insignificant but there is a sharp east-west cline both across the Scandinavia, between Finnic and Germanic-speaking populations, and in continental Europe, between Lithuania (about 40%) and Poland (about 2%). Less sharp but still significant is a drop of N3 frequency from Lithuanians to Belorussians, among the latter its frequency is around 6% to 2% (6% according to our unpublished data, n=100; 2% in Rosser et al. 2000, n=41). In western and southern regions of Europe this haplogroup is practically absent (fig. 6 in the thesis, table 3 in ref. II).

From the archaeological point of view, hg N3 is spread in Europe in the area of comb-ceramic culture. It is not, however, obvious that the spread of the two can be temporarily connected, because STR diversty-based calculations of the time depth of hg N3 among the Finnic-speaking European populations suggests expansion time before-around the end of Pleistocene — (our unpublished data, L. Zhivatovsky personal communication) that is long before the rise of the comb-ceramic culture in the 4th millennium BC. The presence of hg N3 in moderate frequencies (around 15% in average) in Russian population is most logically explained by Finnic genetic substratum in East Europe. The near absence of haplogroup N3 in western and southern Slavs, as well as the historically recorded arrival of Slavonic tribes to East Europe about 15 centuries ago, may well explain this frequency estimate. Broadly, it tells that the expected substratum may comprise, in average, nearly one fifth of paternal lineages of extant Russians. It is likely that this average figure may vary considerably among Russians. Indeed, it is already known that the northern (Pomori) Russians possess hg N3 at much higher frequency (20–43% according to Wells et al. 2001) than Russians in average.

As haplogroup N3 is widely spread both in Siberia and northern Europe, the question of the direction of its spread arises. This is a complex question and should be formulated clearly to avoid misunderstandings. In a direct meaning, it is question from where the expansion of N3, leading to its spread in northern-northeastern Europe and in Siberia, has started. However, there is also a question about the place and time of the origin of haplogroup N, including an even more general problem about the temporal and spatial location of the split between haplogroups N and O — the two sister groups that share M214 biallelic marker, common for the two clades (see fig. 7 in the thesis).

The answer to the second problem is still poorly understood — it could have happened in an area from Southeast Asia to North China, or even in more western areas. An answer to the first question is somewhat more at hand. According to the recent studies (ref. IV, II and references therein), the diversity of hg N3 in eastern Europe, compared to Siberian populations (Yakuts, Buryats) (fig.5, ref. IV), is much higher, which suggests that eastern Europe, rather than Siberia, is a possible origin of the detectable at present expansion of N3 in North Eurasia. However, the closely related hg N2 lineage (sister-clade of N3) is frequent in Siberians, but present only in some Volga-Uralic populations (see table 3 in ref. II). It hints that further back in time, the common ancestor (M231) for N2-P43 and N3-M178 chromosomes, may have arisen rather in Asia than in East Europe. Therefore, it is possible that hg N variation in East Europe and Siberia is testifying about a considerable, deep late Pleistocene/early Holocene share of paternal lineages between populations that later on became ancestors to Uralic and Altaic-speaking people.

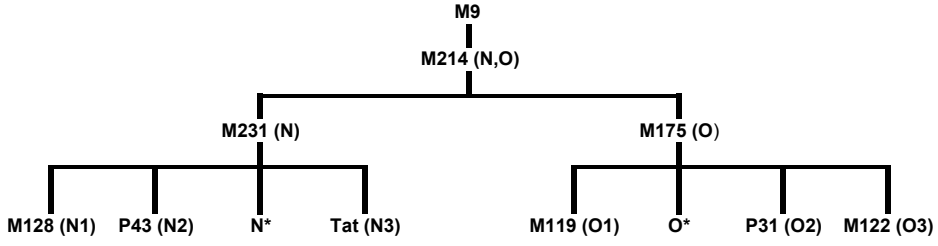

Figure 7. Schematic view of phylogenetic relationship between M214 defined clades.

Besides hg N3, another widely spread and characteristic haplogroup among eastern European populations is R1a (ref.II, IV and referred as haplogroup 3 in ref. IV; see also fig. 2 in the thesis). However, this haplogroup is present also in Central-Asia and India, what means that at its present-day resolution, it is not informative enough to reveal details of its spread: novel informative down-stream mutations are needed to get further insight into the phylogeography of R1a in whole Eurasia.

In Europe, spread of hg R1a is characterized by a gradient, opposite to that for R1b, which reaches particularly high frequencies in western Europe, whereas high frequency of hg R1a is characteristic to Slavic-speaking populations (ref. II, IV) comprising about a half of their paternal gene pool. The spread of this haplogroup can be connected with post-LGM recolonization of Europe from the Periglacial refugium in the territory of the present-day Ukraine (Semino et al. 2000a), particularly, as the highest 49a,f/Taq I haplotypes diversity of haplogroup R1a has been shown in the Ukrainians (Passarino et al. 2001a). However, as alreay mentioned above, there is also another interpretation to the present-day phylogeography of haplogroup R1a, linking it to a much more recent spread of the Kurgan culture in East European steppe belt (Rosser et al. 2000).

Clear longitudinal frequency clines like for Y-chromosomal R1b and R1a are not apparent for widespread "classical" European mtDNA haplogroups which have typically much smoother distribution patterns over whole Europe. However, a deeper phylogenetic dissecting of several pan-European mtDNA variants, such as U4, U5 and H (Tambets 2003; Tambets 2004; Loogväli et al. 2004; Achilli et al. 2004) show that better understanding of mtDNA phylogeograpy can be informative in finding parallels in the spread of the two haploid genetic systems.

6. CONCLUSIONS

1. The refinement in phylogenetic resolution of European-specific Y-chromosomal haplogroup I enabled to distinguish distinct geographic distribution of its sub-clades I1a, I1b and I1c, defined by bi-allelic markers and to establish their phylogeograpy.

2. The phylogeographic distribution patterns, diversity and expansion times indicate that the present spread pattern of sub-clades of haplogroup I reveal postglacial re-colonization of Europe from different refugial areas. According to obtained data, sub-clades I1a, I1c and I1b2 have likely spread from Iberian/Southern France (Franco-Cantabrian refugium) and I1b* has started its spread from the Balkans or East Europe (Periglacial refugium in the Ukraine).

3. The analysis of Y-chromosomal markers showed that paternal lineages of the Saami originate from the source, shared with other northern European populations. The Y-chromosomal haplogroup variation in the Saami population does not reflect severe bottleneck events, presumably responsible for severely reduced selection of mtDNA variants in this population.

4. Y-chromosomal haplogroup N3 is widely distributed over the whole northern Eurasia. The diversity of haplogrop N3 in northern and eastern Europe is much higher than in Siberia, suggesting that the present phylogeographic distribution of this haplogroup is best explained by its initial postglacial expansion from eastern Europe, rather than from Siberia, where its frequency, in some populations, exceeds that among eastern Europeans.

REFERENCES

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F et al. (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. Epub. Sep.20 (2004)

Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A, Santachiara-Benerecetti AS (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. Mol Phylogenet Evol 28:458–72.

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

Armour JA, Anttinen T, May CA, Vega EE, Sajantila A, Kidd JR, Kidd KK, et al (1996) Minisatellite diversity supports a recent African origin for modern humans. Nat Genet 13:154–60

Avise J (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambidge, Massachusetts London

Avise J, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Rub CA, et al (1887) Intraspecific phylogeography: the molecular bridge between population genetics and systematics. Ann. Rev. Ecol. Syst. 18:489–522

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

Bao W, Zhu S, Pandya A, Zerjal T, Xu J, Shu Q, Du R, et al (2000) MSY2: a slowly evolving minisatellite on the human Y chromosome which provides a useful polymorphic marker in Chinese populations. Gene 244:29–33.

Barac L, Pericic M, Klaric IM, Rootsi S, Janicijevic B, Kivisild T, Parik J, et al (2003) Y chromosomal heritage of Croatian population and its island isolates. Eur J Hum Genet 11:535–42.

Barbujani G, Bertorelle G, Chikhi L (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. Am J Hum Genet 62:488–92

Bar-Yosef O (1998) The Natufian Culture in the Levant, Threshold to the Origins of Agriculture. Evolutionary Anthropology 6:159–177

Belledi M, Poloni ES, Casalotti R, Conterio F, Mikerezi I, Tagliavini J, Excoffier L (2000) Maternal and paternal lineages in Albania and the genetic structure of Indo-European populations. Eur J Hum Genet 8:480–6.

Bergen AW, Wang CY, Tsai J, Jefferson K, Dey C, Smith KD, Park SC, et al (1999) An Asian-Native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. Ann Hum Genet 63:63–80.

Blanco P, Shlumukova M, Sargent CA, Jobling MA, Affara N, Hurles ME (2000) Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. J Med Genet 37:752–8.

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. Am J Hum Genet 68:1019–29.

Bosch E, Calafell F, Rosser ZH, Norby S, Lynnerup N, Hurles ME, Jobling MA (2003) High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. Hum Genet 112:353–63. Epub 2003 Feb 20.

Boyd R, Silk JB (1997) How humans evolved. WW Norton, New York

Calafell F, Shuster A, Speed WC, Kidd JR, Kidd KK (1998) Short tandem repeat polymorphism evolution in humans. Eur J Hum Genet 6:38–49.

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–6

Capelli C, Redhead N, Abernethy JK, Gratrix F, Wilson JF, Moen T, Hervig T, et al (2003) A Y chromosome census of the British Isles. Curr Biol 13:979–84.

Capelli C, Wilson JF, Richards M, Stumpf MP, Gratrix F, Oppenheimer S, Underhill P, et al (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. Am J Hum Genet 68:432–43.

Carvajal-Carmona. LG, Soto. JD, N P, Ortziz-Barrientos. D, C. D, Ospina-Duque. J, McCarthy. M, et al (2000) Strong Amerind/White Sex Bias and a Possible Sephardic Contribution among the Founders of a Population in Northwest Colombia. Am. J. Hum. Genet. 67:1287–1295

Carvalho-Silva DR, Santos FR, Rocha J, Pena SD (2001) The phylogeography of Brazilian Y-chromosome lineages. Am J Hum Genet 68:281–6.

Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Fellous M, Purrello M, et al (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. Science 230:1403–6.

Cavalli-Sforza LL, Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. Nat Genet 33 Suppl:266–75

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The History and Geography of Human Genes. Princeton University Press, Princeton

Cavalli-Sforza LL, Minch E (1997) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 61:247–54

Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. Comp Biochem Physiol B Biochem Mol Biol 126:455–76.

Charlesworth B (1996) The evolution of chromosomal sex determination and dosage compensation. Curr Biol 6:149–62.

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–20.

Chikhi L, Destro-Bisol G, Bertorelle 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 U S A 95:9053–8

Chikhi L, Nichols RA, Barbujani G, Beaumont MA (2002) Y genetic data support the Neolithic demic diffusion model. Proc Natl Acad Sci U S A 99:11008–13.

Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, et al (1998) Genetic relationship of populations in China. Proc Natl Acad Sci U S A 95:11763–8

Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, et al (2004) Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet. 114:127–148

Collins FS, Guyer MS, Charkravarti A (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278:1580–1.

Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, Bertranpetit J (2004) Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. Eur J Hum Genet 11:1–10

Cooke HJ, Brown WR, Rappold GA (1985) Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. Nature 317: 687–692

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, et al (2004) Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. Am J Hum Genet 74:1014–1022

de Knijff P, Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, et al (1997) Chromosome Y microsatellites: population genetic and evolutionary aspects. Int J Legal Med 110:134–49

de Knijff P (2000) Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am. J. Hum. Gen. 67: 1055–1061

Dennell R (1983) European economic prehistory: a new approach. Academic Press, London

Derenko MV, Malyarchuk BA, Denisova GA, Dorzhu CM, Karamchakova ON, Luzina FA, lotosh EA, et al (2002) Polymorphism of the Y-Chromosome Diallelic Loci in Ethnic Groups of the Altai-Sayan Region. Russian Journal of Genetics 38:309–314

Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, Bricka R et al. (2004) Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. Human Genetics, Epub. 21. Aug. 2004

Ding YC, Wooding S, Harpending HC, Chi HC, Li HP, Fu YX, Pang JF, et al (2000) Population structure and history in East Asia. Proc Natl Acad Sci U S A 97:14003–14006

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

Dupuy BM, Andreassen R, Flones AG, Tomassen K, Egeland T, Brion M, Carracedo A, et al (2001) Y-chromosome variation in a Norwegian population sample. Forensic Sci Int 117:163–73.

Fernandes S, Huellen K, Goncalves J, Dukal H, Zeisler J, Rajpert De Meyts E, Skakkebaek NE, et al (2002) High frequency of DAZ1/DAZ2 gene deletions in patients with severe oligozoospermia. Mol Hum Reprod 8:286–98.

Fernandes S, Paracchini S, Meyer LH, Floridia G, Tyler-Smith C, Vogt PH (2004) A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in Y haplogroup N. Am J Hum Genet 74:180–7.

Forster P, Kayser M, Meyer E, Roewer L, Pfeiffer H, Benkmann H, Brinkmann B (1998) Phylogenetic resolution of complex mutational features at Y-STR DYS390 in aboriginal Australians and Papuans. Mol Biol Evol 15:1108–14.

Forster P, Rohl A, Lunnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B (2000) A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet 67:182–96.

Francalacci P, Morelli L, Underhill PA, Lillie AS, Passarino G, Useli A, Madeddu R, et al (2003) Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. Am J Phys Anthropol 121:270–9.

Freije D, Helms C, Watson MS, Donis-Keller H (1992) Identification of a second pseudoautosomal region near the Xq and Yq telomeres. Science 258: 1784–1787.
Gibbons A (2003) Physical anthropology and paleoanthropology meeting. First modern remains in Europe. Science 300:894.
Graves JA (1996) Breaking laws and obeying rules. Nat Genet 12:121–2.
Graves JA, Schmidt MM (1992) Mammalian sex chromosomes: design or accident? Curr Opin Genet Dev 2:890–901.
Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, et al (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15:427–41
Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL (2001) Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol 18:1189–203.
Hammer MF, Zegura SL (2002) The Human Y Chromosome Haplogroup Tree: Nomenclature and Phylogeography of its Major Divisions. Annu. Rev. Anthropol. 31:303–321
Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST (1998) Genetic traces of ancient demography. Proc Natl Acad Sci U S A 95:1961–7
Helgason A, Sigureth ardottir S, Nicholson J, Sykes B, Hill EW, Bradley DG, Bosnes V, et al (2000) Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. Am J Hum Genet 67:697–717.
Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature 405: 907–913
Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6:799–803.
Hurles ME, Irven C, Nicholson J, Taylor PG, Santos FR, Loughlin J, Jobling MA, et al (1998) European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. Am J Hum Genet 63:1793–806
Hurles ME, Jobling MA (2003) A singular chromosome. Nat Genet 34:246–7.
Hurles ME, Maund E, Nicholson J, Bosch E, Renfrew C, Sykes BC, Jobling MA (2003) Native American Y chromosomes in Polynesia: the genetic impact of the Polynesian slave trade. Am J Hum Genet 72:1282–7.
Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713
Jegalian K, Page DC (1998) A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. Nature 394:776–80.
Jin HJ, Kwak KD, Hammer MF, Nakahori Y, Shinka T, Lee JW, Jin F, et al (2003) Y-chromosomal DNA haplogroups and their implications for the dual origins of the Koreans. Hum Genet 114:27–35.
Jin L, Su B (2000) Natives or immigrants: modern human origin in east Asia. Nat Rev Genet 1:126–33.
Jin ZB, Huang XL, Nakajima Y, Yukawa N, Osawa M, Takeichi S (2003) Haploid allele mapping of Y-chromosome minisatellite, MSY1 (DYF155S1), to a Japanese population. Leg Med (Tokyo) 5:87–92.
Jobling MA (2001) In the name of the father: surnames and genetics. Trends Genet 17:353–7.

Jobling MA, Bouzekri N, Taylor PG (1998) Hypervariable digital DNA codes for human paternal lineages: MVR-PCR at the Y-specific minisatellite, MSY1 (DYF155S1). Hum Mol Genet 7:643–53.

Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118–24

Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, Gerelsaikhan T, et al (1996) Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. Hum Mol Genet 5:1767–75.

Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. Trends Genet 11:449–56

Jobling MA, Tyler-Smith C (2000) New uses for new haplotypes: the human Y chromosome, disease and selection . Trends Genet 16:356–62

Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612.

Jobling MA, Williams GA, Schiebel GA, Pandya GA, McElreavey GA, Salas GA, Rappold GA, et al (1998) A selective difference between human Y-chromosomal DNA haplotypes. Curr Biol 31:1391–1394

Jovelin F, Berthaud S, Lucotte G (2003) Molecular basis of the TaqI p49a,f polymorphism in the DYS1 locus containing DAZ genes. Molecular Human Reproduction 9: 509–516

Kageyama M, Peyron O, Pinot S, Tarasov P, Guidot J, Joussaume S, Ramstein G (2001) The Last Glacial Maximum climate over Europe and western Siberia: a PMIP comparison between models and data. Climate Dynamics 17:23–43

Kalaydjieva L, Calafell F, Jobling MA, Angelicheva D, de Knijff P, Rosser ZH, Hurles ME, et al (2001) Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. Eur J Hum Genet 9:97–104.

Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, et al (2001) Paternal population history of East Asia: sources, patterns, and microevolutionary processes. Am J Hum Genet 69:615–28.

Karafet TM, Osipova LP, Gubina MA, Posukh OL, Zegura SL, Hammer MF (2002) High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. Hum Biol 74:761–89.

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, et al (1999) Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. Am J Hum Genet 64:817–31

Kayser M, Brauer S, Weiss G, Schiefenhovel W, Underhill PA, Stoneking M (2001) Independent histories of human Y chromosomes from Melanesia and Australia. Am J Hum Genet 68:173–190.

Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schiefenhovel W, Stoneking M (2000a) Melanesian origin of Polynesian Y chromosomes. Curr Biol 10:1237–46.

Kayser M, Kittler R, Ralf A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, et al (2004) A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 74:1183–97.

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Kruger C, et al (2000b) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66:1580–8.

Kharkov VN, Stepanov VA, Borinskaya SA, Kozhekbaeva ZM, Gusar VA, Grechanina EY, Puzyrev VP, et al (2004) Gene Pool Structure of Eastern Ukrainians as Inferred from the Y-Chromosome haploroups. Russian Journal of Genetics 40:326–331

King R, Underhill Pa (2002) Congruent distribution of Neolithic painted pottery and ceramic figurines with Y-chromosome lineages. Antiquity 76:707–714

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, et al (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr Biol 9:1331–1334

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, et al (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet 72:313–332

Klein RG (2003) Whither the Neanderthals? Science 299:1525–1527

Krausz C, Quintana-Murci L, McElreavey K (2000) Prognostic value of Y deletion analysis: what is the clinical prognostic value of Y chromosome microdeletion analysis? Hum Reprod 15:1431–4.

Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, Wilson RK, et al (2001) The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nat Genet 29:279–86.

Kuzniar P, Ploski R (2004) STR data for the power plex-16 loci in a population from Central Poland. Forensic Sci Int 139:261–3.

Kvist L, Ruokonen M, Lumme J & Orell M (1999) The colonization history and present-day population structure of the European great tit (Parus major major). Heredity 82: 495–502.

Kvist L (2000) Phylogeny and phylogeography of European Parids. Dissertation . Oulu University

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 7:447–58

Lahn BT, Page DC (1997) Functional coherence of the human Y chromosome. Science 278:675–80.

Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. Science 286:964–7.

Lahr M, Foley R (1994) Multiple dispersals and modern human origins. Evolutionary Anthropology 3:48–60

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:137–76

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.

Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, et al (2002) The dual origin and Siberian affinities of Native American Y chromosomes. Am J Hum Genet 70:192–206.

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401.

Loogvali EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K et al. (2004) Disuniting Uniformity: A Pied Cladistic Canvas of mtDNA Haplogroup H in Eurasia. Mol Biol Evol. Epub Jul 14, 2004

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, Roseman C, Underhill PA, et al (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Am J Hum Genet 74:532–544

Maca-Meyer N, Sanchez-Velasco P, Flores C, Larruga JM, Gonzales AM, Oterino A, Leyva-Cobian F (2003) Y Chromomosome and Mithocondrial DNA Characterizatin of Pasiegos , A Human Isolate from Cantabria (Spain). Annals of human Genetics 67:327–339

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P et al. (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. Epub Aug 31, 2004

Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. Genetics 156:297–304.

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, et al (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235:1616–22.

Nasidze I, Sarkisian T, Kerimov A, Stoneking M (2003) Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. Hum Genet 112:255–61.

Nebel A, Filon D, Brinkmann B, Majumder PP, Faerman M, Oppenheim A (2001) The Y chromosome pool of jews as part of the genetic landscape of the middle east. Am J Hum Genet 69:1095–112.

Ngo KY, Vergnaud G, Johnsson C, Lucotte G, Weissenbach J (1986) A DNA probe detecting multiple haplotypes of the human Ychromosome. Am. J. Hum. Genet. 38, 407–418.

Ohno S (1967) Sex chromosomes and sex-linked genes. Springer, Berlin

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 10:521–9.

Passarino G, Semino O, Magri C, Al-Zahery N, Benuzzi G, Quintana-Murci L, Andellnovic S, et al (2001a) The 49a,f haplotype 11 is a new marker of the EU19 lineage that traces migrations from northern regions of the Black Sea. Hum Immunol 62:922–32.

Passarino G, Underhill PA, Cavalli-Sforza LL, Semino O, Pes GM, Carru C, Ferrucci L, et al (2001b) Y chromosome binary markers to study the high prevalence of males in Sardinian centenarians and the genetic structure of the Sardinian population. Hum Hered 52:136–9.

Perez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martinez-Arias R, Clarimon J, et al (1999) Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. Am J Hum Genet 65:208–19.

Peyron O, Guidot L, Cheddadi R, Tarasov P, Reille M, de Beaulieu J-L, Bottema S, et al (1998) Climatic reconstractions in Europe for 18 000 yr B. P. from pollen data. Quat. Res. 49:183–196

Piazza A, Rendine S, Minch E, Menozzi P, Mountain J, Cavalli-Sforza LL (1995) Genetics and the origin of European languages. Proc Natl Acad Sci U S A 92:5836–40

Ploski R, Wozniak M, Pawlowski R, Monies DM, Branicki W, Kupiec T, Kloosterman A, et al (2002) Homogeneity and distinctiveness of Polish paternal lineages revealed by Y chromosome microsatellite haplotype analysis. Hum Genet 110:592–600.

Quintana-Murci L, Chaix R, Wells S, Behar D, Sayar H, Scozzari R, Rengo C, *et al* (2004) Where West meets East: The complex mtDNA landscape of the Southwest and Central Asian corridor. Am J Hum Genet 74:827–845

Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ, Ayub Q, et al (2001) Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. Am J Hum Genet 68:537–42.

Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat Genet 23:437–441

Raitio M, Lindroos K, Laukkanen M, Pastinen T, Sistonen P, Sajantila A, Syvanen A (2001) Y-chromosomal SNPs in Finno-Ugric-speaking populations analyzed by minisequencing on microarrays. Genome Res 11:471–82

Renfrew C, Boyle K (2000) Archaeogenetics: DNA and the population prehistory of Europe. (McDonald Institute Monographs). McDonald Institute for Archaeological Research., Cambridge

Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, et al (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet 35:247–51. Epub 2003 Oct 5.

Repping S, Skaletsky H, Lange J, Silber S, Van Der Veen F, Oates RD, Page DC, et al (2002) Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am J Hum Genet 71:906–22.

Repping S, van Daalen SK, Korver CM, Brown LG, Marszalek JD, Gianotten J, Oates RD, et al (2004) A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. Genomics 83:1046–52.

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

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, et al (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59:185–203

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, et al (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 71:1168–74.

Rootsi S, Kivisild T, Tambets K, Adojaan M, Parik J, Reidla M, Metspalu E, et al (2000) On the phylogeographic context of sex-specific genetic markers of Finno-Ugric populations. 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 148–164

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, et al (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in europe. Am J Hum Genet 75:128–37.

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, et al (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67:1526–43.

Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, et al (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423:873–6.

Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, Moral P, Latini V, et al (2001) Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. Hum Immunol 62:871–84.

Seielstad M, Bekele E, Ibrahim M, Toure A, Traore M (1999) A view of modern human origins from Y chromosome microsatellite variation. Genome Res 9:558–67.

Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278–80

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, et al (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74:1023–34.

Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS (1996) A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. Am J Hum Genet 59:964–8

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, et al (2000a) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science 290:1155–9.

Semino O, Passarino G, Quintana-Murci L, Liu A, Beres J, Czeizel A, Santachiara-Benerecetti AS (2000b) MtDNA and Y chromosome polymorphisms in Hungary: inferences from the palaeolithic, neolithic and Uralic influences on the modern Hungarian gene pool. Eur J Hum Genet 8:339–46.

Semino O, Santachiara-Benerecetti A, Falaschi F, Cavalli-Sforza L, Underhill P (2002) Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. Am J Hum Genet 70:265–8

Shen P, Wang F, Underhill PA, Franco C, Yang WH, Roxas A, Sung R, et al (2000) Population genetic implications from sequence variation in four Y chromosome genes. Proc Natl Acad Sci U S A 97:7354–9.

Simmler MC, Royer F, Vergnaud G, Nystrom-Lahti M, Ngo KY, de la Chapelle A, Weissenbach J (1985) Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. Nature 317: 692–697.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, et al (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–37.

Stefan M, Stefanescu G, Gavrila L, Terrenato L, Jobling MA, Malaspina P, Novelletto A (2001) Y chromosome analysis reveals a sharp genetic boundary in the Carpathian region. Eur J Hum Genet 9:27–33.

Stepanov V, A. (2002) Ethnogenomics of population of Siberia and Central Asia (in Russian). Pechatnaja Manufaktura, Tomsk

Stringer C (2000) Coasting out of Africa. Nature 405:24–5, 27

Stringer C (2003) Human evolution: Out of Ethiopia. Nature 423:692–3, 695.
Stringer C, McKie R (1996) African exodus. The origins of modern humanity. Butler & Tanner, Frome and London
Stringer CB, Andrews P (1988) Genetic and fossil evidence for the origin of modern humans. Science 239:1263–8
Su B, Jin L (2000) Natives or immigrants: modern human origin in East Asia. Nature Reviews 1:126–133
Su B, Xiao C, Deka R, Seielstad MT, Kangwanpong D, Xiao J, Lu D, et al (2000) Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. Hum Genet 107:582–90.
Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, et al (1999) Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last ice age. Am J Hum Genet 65:1718–24
Sun C, Skaletsky H, Birren B, Devon K, Tang Z, Silber S, Oates R, et al (1999) An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. Nat Genet 23:429–32.
Tajima A, Hayami M, Tokunaga K, Juji T, Matsuo M, Marzuki S, Omoto K, et al (2004) Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. J Hum Genet 49:187–193
Tajima A, Pan IH, Fucharoen G, Fucharoen S, Matsuo M, Tokunaga K, Juji T, et al (2002) Three major lineages of Asian Y chromosomes: implications for the peopling of east and southeast Asia. Hum Genet 110:80–8.
Taberlet, P., and J. Bouvet. (1994) Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the brown bear (Ursus arctos) in Europe. Proceedings of the Royal Society of London Series B 255:195–200.
Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogvali EL, Tolk HV, et al (2004) The Western and Eastern Roots of the Saami — the Story of Genetic "Outliers" Told by Mitochondrial DNA and Y Chromosomes. Am J Hum Genet 74: 661–82
Tambets K (2004) Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Dissertaiones Biologicae Universitas Tartuensis 92. Tartu University Press, Tartu
Tambets K, Tolk HV, Kivisild T, Metspalu E, Parik J, Reidla M, Voevoda M et al. (2003) Complex Signals for Population Expansions in europe and Beyond. In: Examining the Farming/Language Dispersals Hypothesis. Cambridge University Press: 449–457
Thangaraj K, Singh L, Reddy A, Rao V, Sehgal S, Underhill P, Pierson M, et al (2003) Genetic affinities of the andaman islanders, a vanishing human population. Current Biology 13:86–93
Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, McPherson JD, et al (2001) A physical map of the human Y chromosome. Nature 409:943–5.
Tolk HV, Barac L, Pericic M, Martinovic-Klaric J, Janisievic B, Campbell H, Rudan I, et al. (2001) The evidence of mtDNA haplogroup F in a European population and its ethnohistoric implications. Eur J Hum Gen 9: 717–723.
Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, et al (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. Am J Hum Genet 69:844–852.

Torroni A, Bandelt H-J, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, et al (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62:1137–52

Tyler-Smith C, Oakey RJ, Larin Z, Fisher RB, Crocker M, Affara NA, Ferguson-Smith MA, et al (1993) Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. Nat Genet 5:368–75.

Underhill PA (2003) Inferring Human History: Clues from Y-Chromosome Haplotypes Cold Spring Harbor Symposia on Quantitative Biology. Vol. LXVIII. Cold Spring Harbor Laboratory Press: 487–493

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, et al (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996–1005

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley R, Oefner PJ, et al (2001a) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet. 65:43–62

Underhill PA, Passarino G, Lin AA, Marzuki S, Oefner PJ, Cavalli-Sforza LL, Chambers GK (2001b) Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. Hum Mutat 17:271–80.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, et al (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al (2001) The sequence of the human genome. Science 291:1304–51.

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–7

Villems R, Rootsi S, Tambets K, Adojaan M, Orekhov V, Khusnutdinova E, Yankovsky N (2002) Archaeogenetics of Finno-Ugric speaking populations. In: Julku K (ed) The Roots of Peoples and Languages of Northern Eurasia IV. Societas Historiae Fenno-Ugricae, Oulu: 271–284

Vogt PH (1998) Human chromosome deletions in Yq11, AZF candidate genes and male infertility: history and update. Mol Hum Reprod 4:739–44.

Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Kohn FM, et al (1996) Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. Hum Mol Genet 5:933–43.

Wainscoat JS, Kulozik AE, Ramsay M, Falusi AG, Weatherall DJ (1986) A Taq 1 gamma-globin DNA polymorphism: an African-specific marker. Hum Genet 74:90–2.

Walter RC, Buffler RT, Bruggemann JH, Guillaume MM, Berhe SM, Negassi B, Libsekal Y, et al (2000) Early human occupation of the Red Sea coast of Eritrea during the last interglacial. Nature 405:65–9.

Weale ME, Weiss DA, Jager RF, Bradman N, Thomas MG (2002) Y chromosome evidence for Anglo-Saxon mass migration. Mol Biol Evol 19:1008–21.

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–8.

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, et al (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci U S A 98:10244–9.

Wen B, Xie X, Gao S, Li H, Shi H, Song X, Qian T, et al (2004) Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. Am J Hum Genet 74:856–65.

White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Howell FC (2003) Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature 423:742–7.

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 U S A 98:5078–83.

Wolpoff M (1989) Multiregional evolution: the fossil alternative to Eden. In: Mellars P, Stringer C (eds) The human revolution: Behavioral and biological perspectives on the origins of modern humans. Edinburgh University Press, Edinburgh

Wolpoff MH, Hawks J, Frayer DW, Hunley K (2001) Modern human ancestry at the peripheries: a test of the replacement theory. Science 291:293–7.

Wolpoff MH, Spuhler JN, Smith FH, Radovcic J, Pope G, Frayer DW, Eckhardt R, et al (1988) Modern human origins. Science 241:772–4

YCC (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339–48

Yen PH (1998) A long-range restriction map of deletion interval 6 of the human Y chromosome: a region frequently deleted in azoospermic males. Genomics 54:5–12.

Yu M, Zhang Y, Xue Y, Chen F, Wang Q, Huang X, Wang B, et al (2002) A new haplogroup pattern displayed in Fujian Han in China. J Hum Genet 47:95–8.

Zerjal T, Beckman L, Beckman G, Mikelsaar AV, Krumina A, Kucinskas V, Hurles ME, et al (2001) Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. Mol Biol Evol 18:1077–87.

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, et al (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–83

Zerjal T, Wells R, Yuldasheva N, Ruzibakiev R, Tyler-Smith C (2002) A genetic landscape reshaped by recent events: y-chromosomal insights into central Asia. Am J Hum Genet 71:466–82

Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, et al (2004) The effective mutation rate at y chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet 74:50–61.

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 and the population prehistory of Europe. McDonald Institute for Archaeological Research Monograph Series, Cambridge University, Cambridge, pp 57–79

SUMMARY IN ESTONIAN

Fülogeograafiline ülevaade inimese Y-kromosomaalsest varieeruvusest Euroopa populatsioonides

Kui varasematel aegadel kasutati inimpopulatsioonide varieeruvuse uurimisel valdavalt valkude polümorfisme, siis umbes paarkümmend aastat tagasi sai alguse nn. "DNA ajastu". Seega sai võimalikuks uurida differentseeritult ka haploidselt esinevaid Y kromosoomi ja mitokondriaalset DNAd (mtDNA), samuti laiendada hüppeliselt geneetiliste markerite valikut autosoomsetes kromosoomides. Emaliini kaudu päranduva mtDNA uuringute algus ennetas mõnevõrra isaliinis edasikantava Y kromosoomi varieeruvuse selgitamist, kuid nüüdseks on mõlemad geneetilised süsteemid intensiivse uurimise objektiks. Võib kahtluseta öelda, et need kaks haploidset genoomi on andnud lõviosa populatsioonigeneetika "DNA ajastu" revolutsiooni.

Rekombinatsiooni puudumise tõttu kogunevad fenotüübi suhtes neutraalsed mutatsioonid inimkonna mtDNA ja Y kromosoomi geenitiiki tänu juhuslikele mutatsioonidele indiviidide sugurakkudes ja kanduvad valdavalt stohhastiliselt edasi järgmistele põlvkondadele, pakkudes seega autosoomidega võrreldes lihtsamat genealoogilist ülevaadet liini arengust. Seega, vaatamata asjaolule, et nii Y kromosoom (selle nn NRY osa) kui ka mtDNA sisaldavad arvukalt individuaalseid geene, on nad mõlemad geneetilises mõttes üksiklookused.

Inimese Y-kromosomaalsete liinide geograafilist levikut, mille kujunemises on põhiroll olnud juhuslikul geenitriivil ja populatsioonide demograafilisel ajalool, on laialdaselt kasutatud kaasaegse inimese kujunemise, tema eelajalooliste migratsioonide ja geneetilise varieeruvuse uurimisel.

Käesolevas doktoritöös on põhirõhk asetatud Euroopa populatsioonide Y kromosoomi varieeruvuse ja selle kujunemise teede uurimisele. Töö eesmärgiks oli selgitada Y kromosoomi haplogruppide jaotust erinevates Euroopa regioonides ja populatsioonides ja interpreteerida saadud tulemusi geeniliinide, kuid osalt ka populatsioonide päritolu seisukohast — esmajoones tuvastades neid protsesse juhtinud olulisi geenivooge ruumis ja ajas.

Töö kirjanduslikus osas antakse ülevaade Y kromosoomi struktuurist ning kujunemisest, kirjeldatakse Y kromosoomi kui markersüsteemi spetsiifilisi omadusi, mis võimaldavad tema kasutamist uurimisobjektina mitmetes teadusvaldkondades, samuti käsitletakse Y kromosoomi markerite fülogeneesipuud ja haplogruppide levikumustrit ning seda mõjutavaid tegureid.

Varasemates töödes oli enimuuritud piirkondadeks valdavalt Lõuna- ja Lääne-Euroopa, mistõttu käesolevas uurimistöös pakkusid erilist huvi Põhja-Euroopa, Balkani ja Ida-Euroopa populatsioonid, sealhulgas Volga regiooni populatsioonid, mille kohta andmed seni puudusid või olid väga napid.

Põhja-Euroopas oli meile eriliselt huvipakkuvaks saamide populatsioon, kuna varasemates töödes oli näidatud nende erandlikku positsiooni teiste Euroo-

pa populatsioonide hulgaks. Meie eesmärgiks oli selgitada klassikalistest populatsiooniuuringutest järelduvat Saamide geneetilist eripära põhjusi ja allikaid uute fülogeneetiliste ja fülogeograafiliste teadmiste valguses.

Teiseks meid oluliselt huvitavaks eesmärgiks oli uurida süvenenult haplogrupp I fülogeograafiat, kuivõrd see on ainus Euroopas ulatuslikult levinenud Y kromosoomi klaad, mis on praktiliselt ainuomane Euroopa populatsioonidele.

Töö eksperimentaalne osa põhineb Y kromosoomi bi-alleelsete markerite analüüsil erinevates populatsioonides, vastavalt Põhja-Euroopas ja Baltikumis (Rootsi saamid, rootslased, eestlased, lätlased), Lääne-Euroopas (portugallased, šveitslased, prantslased), Balkani piirkonnas (horvaadid, bosnialased, albaanlased, sloveenid, rumeenlased, moldaavlased ja gagauusid), Kesk- ja Ida-Euroopas (tšehhid, slovakid, ungarlased, poolakad, valgevenelased, ukrainlased, venelased erinevatest piirkondadest), Ida-Euroopa Volga regioonis (udmurdid, komid, marid, mordvalased, tšuvaššid, tatarlased) samuti Euroopaga seonduvatest piirkondades, nagu Kaukaasias (grusiinid, osseedid, armeenlased, nogaid, adõgeed, karatšaid), samuti Lähis-Idas (türklased), Makaroneesias (madeiralased, assorlased ja Rohelise Neeme Saarte elanikud) ning Siberis (jakuudid). Analüüsitud proovid (lisaks kohapeal kogutud eestlaste ja Eesti venelaste proovidele) on saadud koostöös mitmete teadusasutustega erinevatest maadest. Proovide analüüs toimus Tartus doktoritöö autori poolt või tema kaasabil ja juhendamisel. Uuritavate Y kromosoomide haplogrupiline kuuluvus tehti kindlaks peamiselt restriktsioonanalüüsi või polümorfsete positsioonide sekveneerimise teel. Paljude uuritud populatsioonide puhul määrati ka lühikeste kordusjärjestuste alleelipikkuste kombinatsioonid e. haplotüübid, mida kasutati haplogrupisisese varieeruvuse uurimiseks. Andmete analüüsiks kasutati haplogruppide sagedusel põhinevat põhikomponentanalüüsi, mediaanvõrgustike konstrueerimist kordusjärjestuste pikkusvariantide põhjal, diversiteedi arvutust ning ekspansiooniaegade arvutust lühikeste kordusjärjestuste varieeruvuse alusel.

Töö põhilised tulemused on kokkuvõtlikult järgmised:

1. Täiuslikuma fülogeneetilise lahutusastme kasutuselevõtt võimaldas eristada Euroopa-keskse Y-kromosomaalse haplogrupi I alam-klaade I1a, I1b ja I1c, mis on defineeritud kindlate bialleelsete markerite poolt. Selgitati nimetatud alam-klaadide fülogeograafia Euroopas.

2. Haplogrupp I alam-klaadide fülogeograafia, diversiteedi jaotus ruumis ja ekspansiooniajad võimaldavad järeldada, et praegune alam-klaadide esinemissagedus ja muster kajastab Euroopa jääaja-järgset rekoloniseerimist, mis meie andmetele tuginedes lähtus eeldatavasti geograafiliselt erinevaist refuugiumeist. Alam-klaadide I1a, I1c ja I1b2 ekspansioon on alanud tõenäoliselt Ibeeriast (Frankokantaabria refuugium) ja I1b* ekspansioon Balkanilt või Ida-Euroopast (periglatsiaalne refuugium tänapäeva Ukrainas).

3. Saamide Y-kromosomaalsete markerite analüüs näitab, et nende seas levinud isaliinid pärinevad geenitiigist, mis on omane Põhja-Euroopa rahvastele tervikuna. Saamide geneetiline kaugus teistest eurooplastest ei ole tingitud nende erinevast päritolust, vaid seletub geenitriiviga, mis ei kajastu isaliinis päritavas Y kromosoomis, kuid on teravalt väljendunud saamide mtDNA varieeruvuses.

4. Y-kromosomaalse haplogrupi N3 leviala katab praktiliselt kogu Euraasia mandri põhjaosa. Euroopa põhja- ja idaosas on N3 haplogrupi-sisene lühikeste kordusjärjestuste haplotüüpide varieeruvus kõrgem kui uuritud Siberi populatsioonides. Seega on tõenäoline, et hoolimata Siberi populatsioonides esinevast N3 kõrgest sagedusest, on selle haplogrupi praegune fülogeograafia kirjeldatav esmajoones Ida-Euroopast lähtunud postglatsiaalse ekspansioonina.

ACKNOWLEDGEMENTS

The study was carried out in the Estonian Biocentre and the Department of Evolutionary Biology, Institute of Molecular and Cell Biology of the University of Tartu.

I have great pleasure to thank my supervisor Prof. Richard Villems for his guidance and being the driving force in completing the manuscript.

I am very grateful to all co-workers in our laboratory for their support, discussion and healthy criticism during the writing period of the thesis. It was very kind of them to create a pleasant, warm and friendly atmosphere during the years of my intense research work.

I would like to give my special thanks to Kristiina Tambets for her permanent patience, cheerful attitude and practical support in all stages of work.

I am truly thankful to Ildus Kutuev for his help with computer graphics and to Ille Ilpus and Jaan Lind for technical assistance.

I sincerely appreciate my closest old friends for their long-lasting friendship and being available for me in good and bad times.

I want to express my greatest gratitude to all partners from different Institutions, who have been involved in the research by forming the collections of the DNA samples:
Dr. Elza Khusnutdinova, Institute of Biochemistry and Genetics in Ufa; Dr. Elena Balanovska, Research Center for Medical Genetics in Moscow; Prof. Pavao Rudan, Institute for Anthropological Research in Zagreb; Wofgang Scheffrahn, Institute of Anthropology in Zurich; Prof. Maya Simonescu, Institute of Cellular Biology and Pathology "Nicolae Simionescu" in Bucharest; Prof. Antonio Brehm, Human Genetics Laboratory, Center of Macaronesian Studies, University of Madeira; Prof. Jean-Paul Moisan, Laboratoire d´Etude du Polymorphisme de l´ADN, Faculté de Médecine in Nantes; Prof. Vladimir Ferak, Department of Molecular Genetics of Comenius University in Bratislava; Prof. Sandor Füredi, Institute of Forensic Sciences in Budapest; Prof. Lars Beckman, Gotland University in Visby; Dr. Ilia Mikerezi, Department of Biology, Faculty of Natural Sciences of Tirana University and Prof. Astrida Krumina, Medical Academy of Latvia in Riga.

Last but not least, I want to express my deepest gratitude to my family for their love and patience in sharing my life during the long years of research and study.

This research was partly supported by Estonian basic research grant 5574 and European Commission grants ICA1CT20070006 and QLG2-CT-2002-90455.

PUBLICATIONS

CURRICULUM VITAE

Siiri Rootsi

Citizenship:	Estonian
Date and place of birth:	January 13th, 1959, Pärnu, Estonia
Family status:	married, three children
Address:	Institute of Molecular and Cell Biology, University of Tartu and Estonian Biocentre, 23 Riia Street, 51010, Tartu, Estonia
Telephone:	+372 7 375 053
Fax:	+372 7 420 286
E-mail:	sroots@ebc.ee

Education and professional employment

1966–1977	Tartu Seconary School No. 2
1977–1982	University of Tartu, Faculty of Biology and Geography, graduated as zoologist, teacher of biology and chemistry
1985–1987	Teacher of biology in Pärnu Secondary School no. 1
Since 1987	Estonian Biocentre, scientist
1996–2002	PhD student in the Department of Evolutionary Biology, research subject: human Y-chromosomal variation

Scientific work

Starting from 1987 I have been working in Estonian Biocentre and from 1996 to 2002 studying as PhD student in the Department of Evolutionary Biology of the University of Tartu under the supervision of Prof. Richard Villems. I have been involved in projects, associated with the analysis of human Y-chromosomal variation in different Eurasian populations. The results of studies are published in the following articles:

1. **Rootsi, S.**, Magri, C., Kivisild, K., Benuzzi, G., Help, H., Bermisheva, M., Kutuev, I., Barać, L., Peričić, M., Balanovsky, O., Pshenichnov, A., Dion, D., Grobei, M., Zhivotovsky, L. A., 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, J.-P., Chaventre, A., Ferak, V., Füredi, S., Oefner, P. O., Shen, P., Beckman, L., Mikerezi, I., Terzić, R., Primorac, D., Cambon-Thomsen, A., Krumina, A., Torroni, A., Underhill, P. A., Santachiara-Bene-

recetti, A. S., Villems, R., Semino, O. (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *American Journal of Human Genetics* 75, 128–137

2. Tambets, K., **Rootsi, S.,** Kivisild, T., Help, H., Serk, P., Loogväli, E.-L., Tolk H.-V., Reidla, M., Metspalu, E., Pliss, L., Balanovsky, O., Pshenichnov, A., Balanovska, E., Gubina, M., Zhadanov, S., Osipova, L., Damba, L., Voevoda, M., Kutuev, I., Bermisheva, M., Khusnutdinova E., Gusar, V., Grechanina, E., Parik, J., Pennarun, E., Richard, C., Chaventre, A., Moisan, J.-P., Barać, L., Peričić, M., Rudan, P., Terzić, R., Mikerezi, I., Krumina, A., Baumanis, V., Koziel, S., Rickards, O., De Stefano, GF., Anagnou, N., Pappa, K.I., Michalodimitrakis, E., Ferák, V., Füredi, S., Komel, R., Beckman, L., Villems, R. (2004) The western and eastern roots of the Saami — the story of genetic "outliers" told by mtDNA and Y-chromosome. *American Journal of Human Genetics* 74, 661–682.
3. Barać L, Peričić M, Martinović Klaric I, **Rootsi S,** Janicijević B, Kivisild T, Parik J, Rudan I, Villems R, Rudan P (2003) Y chromosomal heritage of Croatian population and its island isolates. *Europen Journal of Human Genetics*. 11, 535–542.
4. Barać L., Peričić M., Martinović Klarić I., Janic´ijević B., Parik J., **Rootsi S**., Rudan P. (2003) Y chromosome STRs in Croatians. *Forensic Science International* 138, 127–133
5. 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. (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *American Journal of Human Genetics*. 72, 313–332.
6. T. Kivisild, **S. Rootsi,** M. Metspalu, E. Metspalu, J. Parik, E. Usanga, S.S. Papiha, S. Mastana, R. Villems.(2003) Genetics of the language and farming spread in India. In: *Examining the farming/language dispersal hypothesis*. Eds. P.Bellwwood and C.Renfrew. McDonald Institute Monographs, Cambridge University, 215–222.
7. Villems, R., **Rootsi, S.,** Tambets, K., Adojaan, M., Orekhov, V., Khusnutdinova, E., Yankovsky, N. (2002) Archaeogenetics of Finno-Ugric-speaking populations. In *The Roots of Peoples and Languages of the Northern Eurasia*. (Societas Historiae Fenno-Ugricae, Gummerus Kirjapaino Oy), 271–284.
8. 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-East European populations. *TRAMES*, 5 (55/50), 1, 59–74.
9. Rosser Z H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D. G., Brede, G., Cooper, G., Corte-Real, H. B. S. M., de Knijff, P., Decorte, R., Dubrova, Y. E.,

Evgrafov, O., Gilissen, A., Glisic, S., Gölge, M., Hill, E. W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S. A., Krumina, A., Kucinskas, V., Lavinha, J., Livshits, L. A., Malaspina, P., Maria, S., McElreavey, K., T., Meitinger, A., Mikelsaar, A.-V., Mitchell, R. J., Nafa, K., Nicholson, J., Nørby, S., Pandya, A., Parik, J., Patsalis, P. C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M. J., Previdere´, C., Roewer, L., **Rootsi, S**., Rubinsztein, D. C., Saillard, J., Santos, F. R., Stefanescu, G., Sykes, B. C., Tolun, A., Villems, R., Tyler-Smith, C., Jobling M. A. (2000)Y-Chromosomal Diversity in Europe Is Clinal and Influenced Primarily by Geography, Rather than by Language. *American Journal of Human Genetics* 67, 1526–1543

10. T. Kivisild, S.S. Papiha, **S. Rootsi,** J. Parik, K. Kaldma, M. Reidla, S. Laos, M. Metspalu,G. Pielberg, M. Adojaan, S.S. Mastana, Y. Wang, M. Glögge, H. Demitras, E. Schnakenberg, G.F. De Stefano, T. Geberhiwot, M. Claustres, R. Villems (2000) An Indian ancestry: a key for the understanding of human diversity in Europe and beyond. In: *Archaeogenetics: DNA and the population prehistory of Europe.* Series of Monographs of the McDonald Institute for Archaeological Fesearch, Cambridge University, 267–276

11. Rootsi, S., Kivisild, T., Tambets, K., Adojaan, M., Parik, J., Reidla, M., Metspalu, E., Laos, S., Tolk, H.-V., Villems, R. (2000) On the Phylogeographic Context of Sex-Specific Genetic Markers of Finno-Ugric Populations. In *The Roots of Peoples and languages of Northern Eurasia II and III.* (University of Tartu, Division of Uralic Languages/Societas Historiae Fenno-Ugricae), 148–164.

12. Villems, R., Adojaan, M., Kivisild, T., Parik, J., Pielberg, G., **Rootsi, S.,** Tambets, K**.**, Tolk, H.-V. (1998) Reconstruction of maternal lineages of Finno-Ugric speaking people and some remarks on their paternal inheritance. In *The Roots of Peoples and languages of Northern Eurasia I.* (Societas Historiae Fenno-Ugricae, Gummerus Kirjapaino Oy), 180–200.

CURRICULUM VITAE

Siiri Rootsi

Kodakondsus:	Eesti
Sünniaeg ja -koht:	13 jaanuar, 1959, Pärnu, Eesti
Perekonnaseis:	abielus, peres kolm last
Aadress:	Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituut, Eesti Biokeskus, Riia mnt. 23, 5010, Tartu, Eesti
Telefon:	+372 7 375 053
Fax:	+372 7 420 286
E-post:	sroots@ebc.ee

Haridus ja erialane teenistuskäik

1966–1977	Tartu 2. Keskkool
1977–1982	Tartu Ülikool, Bioloogia- ja Geograafiateaduskond, Lõpetanud bioloogina ja bioloogi keemia õpetajana, zooloogia eriala
1985–1987	Bioloogia õpetaja Pärnu 1. Keskkoolis
Alates 1987	Eesti Biokeskus, teadur
1996–2002	Evolutsioonilise bioloogia õppetooli doktorant, uurimistöö teema: inimese Y-kromosomaalne varieeruvus

Teadustegevus

Alates 1987 a. olen töötanud Eesti Biokeskuses ja aastatel 1996–2002 Evolutsioonilise bioloogia õppetooli juures doktorandina. Olen olnud seotud mitmete teadusprojektidega, mille eesmärgiks on inimese Y-kromosomaalse varieeruvuse uurimine erinevates Euroopa ja Aasia populatsioonides.