DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
371

HELEN POST

Overview of the phylogeny and phylogeography of the Y-chromosomal haplogroup N in northern Eurasia and case studies of two linguistically exceptional populations of Europe – Hungarians and Kalmyks

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
371

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
371

HELEN POST

Overview of the phylogeny and phylogeography of the Y-chromosomal haplogroup N in northern Eurasia and case studies of two linguistically exceptional populations of Europe – Hungarians and Kalmyks

UNIVERSITY OF TARTU
Press

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

Dissertation was accepted for the commencement of the degree of Doctor of Philosophy in Gene Technology on 17 June, 2020 by the Council of the Institute of Molecular and Cell Biology, Faculty of Science and Technology, University of Tartu.

Supervisors: Richard Villems, Professor of Archaeogenetics, Chair of Evolutionary Biology, Institute of Molecular and Cell Biology, and Research Professor, Institute of Genomics, University of Tartu

Siiri Rootsi, PhD; Senior Research Fellow of Population Genetics, Institute of Genomics, University of Tartu, Estonia

Opponent: Maarten Larmuseau, PhD; Professor, Department of Human Genetics, KU Leuven, University of Leuven, Leuven, Belgium

Commencement: Room No 105, 23B Riia St, Tartu, on 26th of August, 2020, at 10:15

Publication of this thesis is granted by the Institute of Molecular and Cell Biology, University of Tartu.

ISSN 1024-6479
ISBN 978-9949-03-379-9 (print)
ISBN 978-9949-03-380-5 (pdf)

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Tables

Figures

LIST OF ORIGINAL PUBLICATIONS

I

Ilumäe AM, Reidla M, Chukhryaeva M, Järve M, **Post H**, Karmin M, Saag L, Agdzhoyan A, Kushniarevich A, Litvinov S, Ekomasova N, Tambets K, Metspalu E, Khusainova R, B Yunusbayev, Khusnutdinova EK, Osipova LP, Fedorova S, Utevska O, Koshel S, Balanovska E, Behar DM, Balanovsky O, Kivisild T, Underhill PA, Villems R, Rootsi. 2016. **Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families**. *American Journal of Human Genetics* 99(1): 163–173. doi: 10.1016/j.ajhg.2016.05.025

II

Post H, Németh E, Klima L, Flores R, Fehér T, Türk A, Székely G, Sahakyan H, Mondal M, Montinaro F, Karmin M, Saag L, Yunusbayev B, Khusnutdinova EK, Metspalu E, Villems R, Tambets K, Rootsi S. 2019. **Y-chromosomal connection between Hungarians and geographically distant populations of the Ural Mountain region and West Siberia.** *Scientific Reports* 9: 7786. doi: 10.1038/s41598-019-44272-6

III

Balinova N, **Post H**, Kushniarevich A, Flores R, Karmin M, Sahakyan H, Reidla M, Metspalu E, Litvinov S, Dzhaubermezov M, Akhmetova V, Khusainova R, Endicott P, Khusnutdinova E, Orlova K, Bakaeva E, Khomyakova I, Spitsina N, Zinchenko R, Villems R, Rootsi S. 2019. **Y-chromosomal analysis of clan structure of Kalmyks, the only European Mongol people, and their relationship to Oirat-Mongols of Inner Asia**. *European Journal of Human Genetics* 27: 1466–1474.
doi: 10.1038/s41431-019-0399-0

Author's contributions to the listed articles are as follows:

Ref I: designed primers and performed laboratory work by genotyping and haplotyping samples. Participated in manuscript writing.

Ref II: designed primers, performed laboratory work, analysed the genotyping data, reconstructed and dated the phylogenetic tree, interpreted the results, provided some of the figures and co-wrote the manuscript.

Ref III: reconstructed and dated the phylogenetic tree, analysed the data, provided some of the figures and co-wrote the manuscript.

ABBREVIATIONS

aDNA – ancient DNA
BCI – Bayesian credible interval
chrX – X chromosome
chrY – Y chromosome
CI – confidence interval
hg – haplogroup
HG – hunter-gatherers
IBD – Identical by descent
ISOGG – International Society of Genetic Genealogy
Kb – kilobase
ky(a) – kilo years (ago)
LRS – long read sequencing
Mb – megabase
MSY – male-specific region
MYA – million years ago
mtDNA – mitochondrial DNA
NGS – Next Generation Sequencing
N_e – effective population size
PAR – pseudoautosomal region
RFLP – restriction fragment length polymorphism
SNP/SNV – single nucleotide polymorphism/variant
STR – short tandem repeat
SRS – short read sequencing
SRY – sex-determining locus Y
TMRCA – Time to the Most Recent Common Ancestor
VUR – Volga-Ural region (Volga-Uralic)
YCC – the Y Chromosome Consortium
Y-DNA – Y- chromosomal DNA

1. INTRODUCTION

Genetic data has been used for decades to quench the human thirst for knowledge about the processes that have affected and moulded the demographic history of populations. Throughout the years the field of population genetics has undergone major changes involving both, types of markers used and methodological/technological advancements. In the beginning research relied on classical markers, more specifically on polymorphisms of different proteins, to study human variation. In the 1980's the field moved on to examining polymorphisms of haploid uniparental DNA markers: mitochondrial (mtDNA) and Y-chromosomal DNA (Y-DNA). The last 10–15 years have been significant in the growth of population genetic advancement due to rapid development of sequencing technologies. Today Next Generation Sequencing (NGS) provides a higher phylogenetic resolution along with a set of unbiased markers. MtDNA and Y-DNA only explain demographic processes of maternal and paternal lineages, respectively. NGS has brought whole genome sequences and autosomal markers that contain more information about the genetic history of populations. Although modern genomic data is instrumental in population genetics and contains a lot of information about the past, it may not give a complete picture of some aspects like time and geographic origins. NGS helps to overcome these limitations as well as enables to sequence the DNA of already deceased individuals called ancient DNA (aDNA) possible. Ancient DNA gives more precise information about the timeframe, geographic origin and can uncover uniparental lineages that have died out.

Haploid, nonrecombining region of Y chromosome passes on through the paternal line and does not undergo a meiotic recombination which means that lineages can be traced back to the most recent common ancestor (TMRCA). Re-sequencing the entire Y chromosome (chrY) thanks to NGS has improved the research of male lineages and their history as the process unveils a lot of markers which can be used in re-constructing a refined phylogeny. Furthermore, these variants and detailed phylogeny makes genotyping and subsequent phylogeographic analysis more time- and cost-efficient. This approach has greatly benefited the global study of chrY variation and has also helped in researching individual populations on a higher resolution level, providing information about specific differences in variation and distinction between populations.

Present thesis focuses on the population genetic aspects of chrY by providing insight into the specific characteristics of the male chromosome, variation and geographic patterns of paternal lineages in north-eastern Eurasia and central regions of Asia. Second part of the dissertation expands the knowledge about the widespread chrY haplogroup (hg) N by utilizing high coverage full sequences of chrY in refining the phylogenetic tree and describing the geographic dispersion of the sub-clades. Similar approach is used in two case studies. First of which involves Uralic-speaking Hungarian population

and hg N3a4 as a genetic link to their putative geographic origin around southern Urals and West Siberia. To achieve a clearer understanding, the phylogeny of aforementioned hg is for the first time re-constructed using Hungarian full chrY sequences combined with samples from other Uralic speakers and populations living near the Uralic mountains. This phylogenetic tree is applied in describing the spread of N3a4 sub-clades in populations of wide geographical range. The second case study analyses the paternal genetic link between the Kalmyk people, the only Mongolic speaking population in Europe, and ethnically related groups inhabiting Mongolia, Kyrgyzstan and China. Correspondingly to the first case study, a phylogeny of the most common hg C3 in Kalmyks is re-constructed. In addition, the genetic profile of paternal lineages in Kalmyks and other Oirats is also revealed.

2. LITERATURE OVERVIEW

The literature overview briefly describes the human Y chromosome including its evolution and characteristics. Additionally, it gives a short overview about whole Y chromosome sequencing, mutation rate estimations and information on Y-chromosomal haplogroups, populations of interest and examples of ancient Y-DNA research of the studied region.

2.1. The features of human Y chromosome

2.1.1. Evolution of human Y chromosome

Human sex chromosomes Y (chrY) and X (chrX) derived from a pair of autosomes (Ohno 1966; Ye et al. 2018) and originated in eutherian mammals around 200–300 million years ago (MYA) (Bachtrog 2013; Lahn and Page 1999; Ross et al. 2006). Both chromosomes have evolved independently in multiple groups of animals including birds, amphibians, reptilians (snakes) and fish (Bachtrog 2013; Charlesworth 1991; Hughes and Rozen 2012). The differentiation between X-Y first began when proto-Y chromosome acquired a sex-determining locus (SRY) (Hughes and Page 2015; Lahn and Page 1999; Trombetta and Cruciani 2017). After that chrY went through multiple inversions which supressed the recombination between the two sex chromosomes (Bellott et al. 2014; Hughes and Page 2015; Hughes and Rozen 2012; Lahn and Page 1999). During chrY evolution crossing over was stifled five times in five separate chromosomal regions (strata) which then went through gene loss and deletions individually (Hughes et al. 2012; Lahn and Page 1999; Ross et al. 2006). Because of that chrY has considerably degenerated both in size and genetic content, compared to chrX (Hughes et al. 2005, 2010, 2012; Hughes and Page 2015; Rice 1996; Skaletsky et al. 2003). Most genes on chrY have become non-functional due to absence of recombination and accumulation of mutations (mut) (Bellott et al. 2014; Skaletsky et al. 2003). Only around 3% of the ancestral genes has been retained in the male-specific region (MSY) of chrY (Hughes et al. 2012; Skaletsky et al. 2003). In contrast, 98% of the ancestral genes have survived in chrX (Bellott et al. 2014; Mueller et al. 2013). However, the dosage difference between chromosomes Y and X is balanced by a collection of dosage-sensitive genes on chrY and the inactivation of one homologue on chrX (Charlesworth 1978; Graves 1995; Lahn et al. 2001). The decay of chrY has slowed down in the last 25 million years (Bellott et al. 2014; Hughes et al. 2005, 2010, 2012).

2.1.2. Characteristics and sequence classes of human Y chromosome

Y chromosome is haploid with a sequence of around 57 Mb (Ensembl GRCh38 release 99). Its main biological functions include testis determination and spermatogenesis. The chromosome consists of the male-specific region (Rozen et al. 2003; Skaletsky et al. 2003) flanked by pseudoautosomal regions (PAR1 and PAR2) (Figure 1) (Hughes and Rozen 2012; Jobling and Tyler-Smith 2003, 2017; Lahn and Page 1999; Skaletsky et al. 2003).

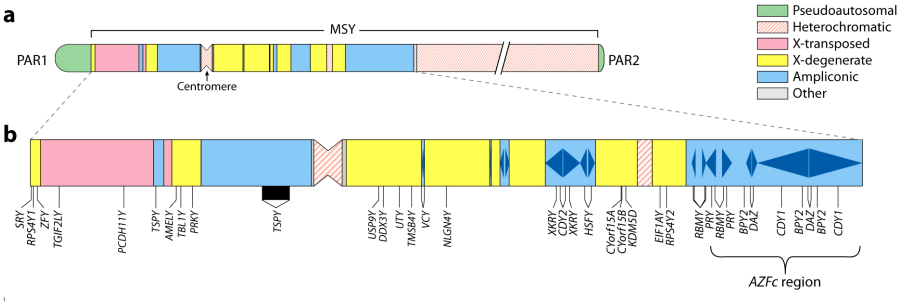

Figure 1. Schematic depiction of the human Y chromosome.
(a) and the male specific region (MSY) with protein coding genes displayed (b). Colours correspond to different sequence classes. Pairs of triangles represent palindromes and arms belonging to the same palindrome are shown as adjacent, opposite triangles. Modified from Figure 1 (Hughes and Rozen 2012) with permission from Annual Reviews, Inc.

MSY is biologically crucial as it carries the *SRY* gene that transcribes a factor important in triggering male development and repressing the development of female reproductive organs (Harley et al. 1992; Ye et al. 2018). MSY contains heterochromatic sequences and three classes of euchromatic sequences – X-degenerate, X-transposed and ampliconic regions (Bachtrog 2013; Hughes and Rozen 2012; Skaletsky et al. 2003). Heterochromatic parts are the centromere, an island of heterochromatin that cuts off the euchromatin, and a significant heterochromatic block (~40 Mb) that comprises of more than 3000 tandem repeats and makes up most of the long arm (Hughes and Rozen 2012; Skaletsky et al. 2003).

The euchromatic region of MSY is approximately 23 Mb (Skaletsky et al. 2003). It contains the human specific X-transposed regions, that have the lowest gene density of the three classes, containing only 2 genes (*TGIF2LY*, *PCDH11Y*) (Bachtrog 2013; Hughes et al. 2010; Hughes and Rozen 2012; Skaletsky et al. 2003). Furthermore, these regions have the highest density of interspersed repeat elements in the euchromatin, around 36% of X-transposed sequences are long interspersed nuclear element 1 (LINE1) elements (Skaletsky et al. 2003). X-transposed blocks are ~99% identical to Xq21 (a region on the long arm of chrX) and both genes have homologues in Xq21 (Skaletsky et al.,

2003). X-transposed regions were formed 3–4 MYA because of an extensive X to Y transposition and later split into two separate segments by an inversion event within MSY short arm (Mumm et al. 1997; Schwartz et al. 1998; Skaletsky et al. 2003).

Most of the MSY euchromatic portions (~18 Mb) are made up of X-degenerate and ampliconic regions (Bachtrog 2013; Hughes and Rozen 2012; Skaletsky et al. 2003). X-degenerate sequences are 8.6 megabase-long decayed versions of the chrX (Hughes and Rozen 2012; Skaletsky et al. 2003). X-degenerate region consists of 27 single-copy genes or pseudogenes that are homologues of different X-linked genes and exhibit up to 96% sequence identity to them. X-degenerate sequences encode all together 16 proteins or protein families out of 27. Twelve genes are expressed ubiquitously, and such widely expressed genes seem to only be present in X-degenerate segments (Hughes and Rozen 2012; Skaletsky et al. 2003).

Seven repetitive ampliconic segments, all together 10.2 Mb, are scattered all over the euchromatic long arm. The sequences in these regions are arranged into inverted repeats, tandem arrays and 8 large palindromes that are 5.7 Mb in length (25% of all chr Y euchromatic sequences) (Skaletsky et al. 2003). A big portion of these segments (60%) exhibit a similarity up to 99.9% with other sequences in MSY (Hughes and Rozen 2012; Skaletsky et al. 2003). Ampliconic regions have the highest density of both coding and non-coding genes among the three classes. There are 9 MSY-specific protein-coding gene families encompassing 60 genes and 75 non-coding transcription units. All except one of the 9 gene families are situated in either palindromes (P1, P2, P3, P4, P5 and P8) or an inverted repeat (IR2) (Skaletsky et al. 2003; Ye et al. 2018). The coding gene families are mainly or exclusively expressed in the testes (Bhowmick et al. 2007; Hughes and Rozen 2012; Skaletsky et al. 2003).

ChrY PAR1 makes up 2.6 Mb of the short-arm tip. In contrast, PAR 2 comprises only around 320 Kb of the tip of the long arm (Mangs and Morris 2007; Ross et al. 2006). PAR regions are the only parts of the chromosome that recombine with X chromosome PAR during male meiosis (Bachtrog 2013; Hughes and Page 2015; Hughes and Rozen 2012; Otto et al. 2011; Skaletsky et al. 2003).

The lack of recombination with chrX in MSY (Bachtrog 2013; Charlesworth 1991; Jobling and Tyler-Smith 2003; Lahn and Page 1999; Teitz et al. 2018) drives the accumulation of deleterious mutations (Muller's ratchet), selection against deleterious mutations (background selection) and genetic hitchhiking where fixation of beneficial mutations will concurrently fix the deleterious mutations they are linked with (Bachtrog 2008, 2013; Filatov et al. 2000; Ye et al. 2018). This coupled with haploidy and the fact that chrY is only present in males lowers the effective population size (N_e) of chrY and therefore it is more affected by genetic drift (Charlesworth and Charlesworth 2000; Filatov et al. 2000; Jobling and Tyler-Smith 2003; Ye et al. 2018). When estimating the sex ratio as 1:1 the population would contain four copies of each autosome, 3 copies of chrX and only one copy of chrY.

Therefore, the N_e of chrY is around one third of the chrX effective population size and one fourth of the N_e of autosomes (Jobling and Tyler-Smith 2003). Although N_e of chrY is comparable to the N_e of mitochondrial DNA (mtDNA) considered the maternal counterpart of chrY, a research (Karmin et al. 2015) has shown estimates of female N_e that are consistently two times higher than male N_e estimates. In addition, many populations display a considerable bottleneck in the N_e of chrY around 8–4 thousand years ago (kya) that is absent from mtDNA. The difference might have occurred due to cultural and sex-specific variance in the number of offspring (Karmin et al. 2015). However, it has also been proposed that the Neolithic bottleneck of chrY might have been a result of competition/wars between patrilineal kin groups (clans) (Zeng et al. 2018).

Although MSY does not recombine, it has been discovered to undergo inter-chromosomal gene conversions which regulate the genetic diversity of portions of MSY called gene conversion hotspots (Cruciani et al. 2010; Niederstätter et al. 2013; Rosser et al. 2009; Trombetta et al. 2010, 2014, 2016, 2017). It has been indicated that these hotspots are not completely Y-linked and that some regions of MSY can swap variants with chrX, autosomes, and the rest of chrY by gene conversion while other parts remain genetically isolated.

2.2. Y-chromosomal whole sequencing – a novel methodology in Y-chromosomal studies

The paucity of MSY recombination means that haplotypes change only due to mutations that accumulate through time (Jobling and Tyler-Smith 2003). Accumulated MSY variants have been used to research the phylogeny and phylogeography of male lineages. Early studies in this field identified only a limited number of single nucleotide variants in modest number of samples that used these variants for genotyping larger sample sets (Cinnioğlu et al. 2004; Cruciani et al. 2004; Karafet et al. 2002; Rosser et al. 2000; Semino et al. 2000; Underhill et al. 2000; Zerjal et al. 1997). However, this limitation caused a bias as other variants in the larger set were left unrecorded (Jobling and Tyler-Smith 2003, 2017).

Sequencing is the most systematic way to characterize the variation of Y chromosome. At the beginning chrY sequences were assembled by mapping and sequencing bacterial artificial chromosome clones (BAC) that contained parts of chr Y from the same individual (Skaletsky et al. 2003; Tilford et al. 2001). However, this approach was too costly and extremely time-consuming for broader application.

Next Generation Sequencing (NGS) (Margulies et al. 2005), including Y chromosome short-read sequencing (SRS), started to pick up in 2010 with the publication of 1000 Genomes Project (Altshuler et al. 2010).

SRS starts with library preparation, amplification and subsequent sequencing. The outcome of this process is multiple short sequencing raw reads (35–300 bp). Different SRS platforms vary considerably by having biases and errors unique to them (Goodwin et al. 2016; Pfeifer 2017). Compared to Sanger sequencing error rates of SRS platforms are higher (~0.1–15%) (Goodwin et al. 2016). Those biases/errors affect correct read alignment and subsequently variant calling and genotyping, meaning that reads must be pre-processed, and quality controlled before alignment to a reference genome. Mapping is highly accurate where the sequence is unique and there are many specific and intricate alignment algorithms to handle the varying aspects of different SRS platforms (Jobling and Tyler-Smith 2017; Pfeifer 2017). However, in some regions of the genome correct alignment is very difficult. The mapping of the Y chromosome is especially challenging and makes the SRS sequencing of the whole chromosome is virtually impossible. For example, in the case of repetitive and low-complexity regions (e.g. heterochromatic parts and X-transposed region) the reads are mapped equally well to many sections of the sequence as the short length of the reads is not able to cover the extent of the repetitive regions (Jobling and Tyler-Smith 2017; Pfeifer 2017). The outcome is an ambiguous alignment that may cause errors in the variant calling. After mapping the alignment is further refined and regions where at least one sample diverges from the reference are determined and genotyping (identification of alleles) is completed. False positives caused by sequencing errors and faulty alignment are then removed by filtering (Pfeifer 2017). The problem of sequencing complexity and lengthy repeats can be overcome with long read sequencing (LRS) where the read length is >10 kilobases on average (Lee et al. 2019; Mantere et al. 2019; Nakano et al. 2017) and with the passage of time LRS is being used more and more, especially in medical research. However, compared to short-read sequencing LRS is costly and therefore still not as widespread in phylogenetic studies.

Important aspect of SRS is coverage (number of unique reads covering the reference). Coverage is crucial in discovering new variants, as low coverage may hinder the detection of rare variants that are only present in a few individuals. This affects the consequent phylogeny by artefactual shortening of the branches (Jobling and Tyler-Smith 2017). Coverage and sequence length largely depend on the platform. Also uncovered sequences between individuals are scattered unevenly. Error patterns unique to separate platforms must be considered in merging data from different origins. However, getting the high-quality overlapping regions from distinct platforms often results in shorter sequences.

Although SRS has its challenges this technology has revolutionized how phylogenetic research is carried out. Around 10Mb of discontinuous unique regions of MSY can still be sequenced using SRS technology (Jobling and Tyler-Smith 2017; Poznik et al. 2013) and the amount of studies (Batini et al. 2015; Francalacci et al. 2013; Hallast et al. 2014; Karmin et al. 2015; Lippold et al. 2014; Pinotti et al. 2019; Poznik et al. 2013, 2016; Rootsi et al. 2013;

Underhill et al. 2015; Wei, Ayub, Chen, et al. 2013; Wei, Ayub, Xue, et al. 2013; Wei, Wang, et al. 2018), that use Y-chromosomal "whole" sequences in the ascertainment of precise and dated phylogenies, has been on the rise ever since the development of SRS.

2.3. Y chromosome mutation rates and The Most Recent Common Ancestor

2.3.1. Mutation rate estimations

Biological variation depends on mutations that are the driving force behind evolution. Although mutations occur both in somatic and germ line cells, only germ line mutations carry on to the next generation. Mutations are mostly the only cause of variation in the MSY and because of that MSY is used in finding mutation rates and estimating the Time to the Most Recent Common Ancestor (TMRCA).

Y-chromosomal mutation rate estimates can be found using one of the three predominant approaches (Balanovsky 2017; Francalacci et al. 2016). First of which applies a known pedigree where the mutations are counted on the genealogical line (e.g. father to son). Then divided by generations or years to obtain a genealogical rate of per-generation or per-year, respectively. Second one, called the evolutionary (or calibration) method (Poznik et al. 2013), is acquired when calibrated molecular clock is applied to a historical event with a known date. The third approach applies ancient DNA (aDNA) (Balanovsky 2017; Francalacci et al. 2016; Kivisild 2017; Rasmussen et al. 2010). This method considers that the aDNA sample has collected fewer mutations compared to modern samples. These assumed missing mutations are then counted and divided by the age of the sample, usually estimated by radiocarbon dating.

ChrY binary markers, often used in forensic analyses, medical studies, evolutionary and phylogeographic research, are single nucleotide polymorphisms (variants) or SNPs (SNV) that designate chrYs to haplogroups/clades. Multiallelic short tandem repeats (Y-STRs) (~2–6 bp) determine haplotypes.

Both STRs and SNPs are applied in chrY mutation rate estimations (e.g. in Tables 1 and 2) and subsequent dating of TMRCA. Before sequencing era, a limited number of biallelic SNPs have been described, which restricted the usage of SNPs in mutation rate estimations. Therefore, for a long period of time, abundant and fast-mutating STRs were used in mutation rate estimates.

There are about 700,000 STR loci (Willems et al. 2014) on the whole genome and no less than 4700 STRs on chrY with mutation rates that are extremely variable (Ballantyne et al. 2010; https://www.ensembl.org). The variability of STRs is connected to the size of the motif, major allele length and whether they reside in the coding or the non-coding region (Willems et al. 2014).

Although the amount of Y-STRs in the genome is substantial, the research community often relies on restricted number of STRs that are incorporated into commercially available kits. Sets such as Applied Biosystem's Yfiler with 17 STRs and Promega's PowerPlex Y23 with 23 STRs are often used, in particular because they became standards in forensic databases. Throughout the years, mostly genealogical STR-based mutation rates have been published (e.g. in Table 1). Genealogical rate estimates based on mutation counting in father-son pairs have resulted in ~2–5 x 10^{-3} locus/generation with some differences depending on the number of loci, sample sizes, populations and STR sets used (Boattini et al. 2016, 2019; Bugoye et al. 2018; Burgarella and Navascués 2011; Claerhout et al. 2018; Čokić et al. 2019; Da Fré et al. 2015; Ge et al. 2009; Goedbloed et al. 2009; Heyer et al. 1997; Hohoff et al. 2007; Sánchez-Diz et al. 2008; Turrina et al. 2006, 2015; Wang and Li 2015; Yang et al. 2018). Only one study (Zhivotovsky et al. 2004) estimated the STR mutation rate by using the evolutionary method. This study analysed haplotype variation among Maoris of New Zealand using their migration to the island as a calibration point. Additionally, they analysed calibration in European Roma and compared Y-STR variation with autosomal STRs. Then the average of the three mutation rates was found.

The contrast between genealogical and evolutionary rates makes it difficult to choose which one to apply when assessing TMRCA. The difference in rates is thought to be due to genetic drift that eliminates haplotypes and recurrent mutations, that diminish genetic diversity (Balanovsky 2017; Hallast et al. 2014; Wei, Ayub, Chen, et al. 2013; Wei, Ayub, Xue, et al. 2013; Zhivotovsky et al. 2006).

Table 1. The Y-chromosomal mutation rate estimations based on STRs.
Modified from Table 2 (Balanovsky 2017) with permission from Springer Nature.

Study reference	Mutation rate (locus/generation 10^{-3})	STR panel*	Approach
Boattini *et al*. 2019	**3.38** (95%CI: 2.57-4.23)	Yfiler	Genealogical
	3.96 (95%CI: 3.18-4.79)	PowerPlex Y23	
Čokić *et al*. 2019	**3.72** (95%CI: 2.68-4.99)	PowerPlex Y23	
Bugoye *et al*. 2018	**2.35** (95%CI: 6.41 x 10^{-4}-6.013)	Yfiler	
Claerhout *et al*. 2018	**5.03** (95%CI: 4.67-5.40)	4 Multiplex kits	
Yang *et al*. 2018	**3.4** (95%CI: 2.5-4.5)	AGCU Y24 Plus	
Boattini *et al*. 2016	**3.254** (95%CI: 2.128-4.506)	Yfiler	
Wang *et al*. 2016	**2.6** (95%CI: 1.9-3.5)	Yfiler	
Da Fré *et al*. 2015	**3.768** (95%CI: 3.542-3.944)	PowerPlex Y23	
Turrina *et al*. 2014	**3.38** (95%CI: 1.36-6.95)	PowerPlex Y23	
Burgarella and Navascués 2011	**2.7**	Yfiler	
	3.89	PowerPlex Y23	
Ge *at al*. 2009	**2.1** (95%CI: 1.7-2.5)	Yfiler	
Goedbloed *et al*. 2009	**2.5** (95%CI: 1.6–3.4)	Yfiler	
Zhivotovsky *et al*. 2004;	**6.9 × 10^{-4}** (95%CI: 5.6-8.2 x 10^{-4})	8 STRs	Evolutionary

* Yfiler and PowerPlex Y23 rates if possible

Due to developments in sequencing technologies during the past decade, the usage of SNPs in the mutation rate estimation has come to the forefront (Table 2). However, because of the limitations in SRS sequencing, estimates essentially rely on the X-degenerate region.

The earliest work where SNP genealogical mutation rate was used was published in 2009 where Xue et al (2009) sequenced the chrYs of two individuals from a Chinese pedigree separated by 13 generations. The most powerful such study (Helgason et al. 2015) used over 750 Icelandic males grouped into 274 patrilines to estimate the point mutation rate for MSY. Also, 9 Kazakh men from the same patrilineal clan were used to obtain a genealogical mutation rate estimation for chrY (Balanovsky et al. 2015). All three rates are similar with overlapping confidence intervals (CI) (Table 2).

One of the first studies using evolutionary approach with SNPs was done in 2013 where Poznik et al. (2013) sequenced the complete chrYs of 69 males. To estimate the mutation rate, two Native American Mayas belonging to chrY hg Q lineages and peopling of the Americas were used (commonly accepted archaeological date 15 kya). The evolutionary mutation rate estimates from limited number of studies remain between 0.65–0.82 x 10^{-9} base/year (Table 2) (Francalacci et al. 2013; Poznik et al. 2013) with CIs that tend not to overlap. This demonstrates the difficulties in combining historical population events and phylogenetics.

The most recent approach that incorporates aDNA for mutation rate estimation of chrY involves counting mutations between ancient and modern samples combined with a reliable radiocarbon age of the ancient sample. One of the first studies (Fu et al. 2014) used a high-quality sequence of an Ust'-Ishim male (~45 kya) from Siberia and a set of modern samples. Another study (Karmin et al. 2015) combined chrYs from two ancient samples – Anzick from North America (12.6 kya) (Rasmussen et al. 2010) and Paleo-Eskimo Saqqaq (4 kya) (Rasmussen et al. 2014) – with contemporary samples. Karmin et al counted only transversions as described in Rasmussen et al. (2014) and merged the two rate estimations attaining a final aDNA-based mutation rate. CIs of all of three aDNA based rate estimates presented in Table 2 overlap with each other.

Table 2. The Y-chromosomal mutation rate estimations based on SNPs.
Modified from Table 1 (Balanovsky 2017) with permission from Springer Nature.

Study reference	Mutation rate (base/year 10^{-9})	Sequence length (Mb)	Approach
Xue *et al* . 2009	**1.0** (95%CI: 0.3-2.5)	10.15	Genealogical
Helgason *et al* . 2015	**0.89** (95%CI: 0.80–0.99)	8.96	
Balanovsky *et al* . 2015	**0.78** (95%CI: 0.62-0.94)	9.97	
Poznik *et al* . 2013	**0.82** (95%CI: 0.72-0.92)	9.99	Evolutionary
Francalacci *et al* . 2013*	**0.65** (95%CI: 0.62-0.68)	8.97	
Fu *et al* . 2014	**0.76** (95%CI: 0.67-0.86)	1.8	Based on aDNA
Karmin *et al* . 2015	**0.74** (95%CI: 0.63-0.95)	8.8	
Trombetta *et al* . 2015	**0.716** (95CI%: 0.619-0.815)	1.5	

*high coverage (avg 17.2x)

2.3.2. Time to the Most Recent Common Ancestor

The time estimate of the Most Recent Common Ancestor substantially depends on the markers and the calculational approach used in mutation rate estimation Before SRS, TMRCA estimations were mostly based on STRs (Hammer et al. 1998; Pritchard et al. 1999; Shi et al. 2010; Wilson and Balding 1998) but also on the restricted resequencing of chrY e.g. single genes or a small selected region (Hammer 1995; Kivisild et al. 2003; Tang et al. 2002; Thomson et al. 2000; Underhill et al. 1997). The range of dates obtained for the chrY global TMRCA by different studies was extremely broad ranging from 30 kya to 190 kya and often having wide CIs as well. It has been shown that STR based rates (compared to SNPs) mostly perform inadequately when applied to TMRCA estimation (Hallast et al. 2014; Karmin et al. 2015; Wei, Ayub, Xue, et al. 2013). However, it has been suggested that the evolutionary rate is more suitable to estimate the age of older nodes while the genealogical approach is better for younger nodes (Hallast et al. 2014; Karmin et al. 2015).

SRS and the subsequent SNP-based dating has made a considerable impact on equating the coalescence age of male and female lineages. Poznik et al. (2013) calculated the estimate of chrY TMRCA to be around 138 kya (95%CI: 120–156 kya), roughly equal to 124 kya (95%CI: 99–148 kya) for mtDNA. An older age of around 200 kya was also proposed by Francalacci et al. (2013), although this study used a mutation rate obtained with low-coverage data from the Sardinian population. Thus, the difference between the two estimates may stem from different sample sets with disparate sequencing coverage. The discovery of a deep and rare A0 lineage, that was formerly considered basal, shifted the Y-chromosomal TMRCA to ~200 kya (Scozzari et al. 2014) which differs from the previous estimation of 142 kya (Cruciani, Trombetta, Massaia, et al. 2011).

The TMRCA for a tree rooted with A00 lineage was dated to ~338 kya (95%CI: 237–581 kya) (Mendez et al. 2013). However, this estimate was originally attained by using an autosomal mutation rate adapted for chrY. A00 lineage and Neanderthal sample from El Sidrón (Spain) were used to estimate the divergence of the Neandertal and modern human chrYs (Mendez et al. 2016). Furthermore, ancient Ust'-Ishim individual (45 kya) (Fu et al. 2014) was used to re-evaluate the maximum likelihood TMRCA for modern humans getting an estimate of of ~275 kya (Mendez et al. 2016). This result is roughly between two previously acquired MRCA estimations. One of those (254 kya) was based on the A00 lineage and mutation rate calibrated with the Ust'-Ishim sample (Karmin et al. 2015). The other used two ancient individuals – Ust'-Ishim and Loschbour (8 kya, Luxembourg) (Lazaridis et al. 2014) – and resulted in an estimate of 291 kya (95%CI: 253–343 kya).

2.4. Main Y-chromosomal haplogroups and their phylogeny

Due to their characteristics, all modern chrYs coalesce back in time to one ancestral sequence. MSY changes due to sequential accumulation of mutations that pass on from father to son. That inheritance mode enables the construction of exclusive phylogenetic trees based on binary SNPs. Phylogenies built in such manner are then often applied in a wider phylogeographic research to elucidate the demographic processes (Avise et al. 1987). SRS technologies have substantially simplified the construction process. However, there are downsides in applying descriptive tree-based approaches. Population genetics perceives phylogenetic trees as stochastic consequences of population genetic processes (Nielsen and Beaumont 2009). Meaning that distribution of trees can be drastically different even if populations have the same demographic histories. The randomness of the trees depends on the amount of offspring an individual has and the segregation of alleles (Nielsen and Beaumont 2009). Additionally, owing to the absence of recombination, males that do not have male descendants will not pass on any information on their chrY. The outcome is that there are lineages that have died out and are not represented in modern phylogenies. Single loci like chrY (or mtDNA) must be used with care when interpreting and salvaging population history as these will not show all the processes of the demographic past. Still, using correct assumptions, a global phylogeny of modern chrYs illustrates the divergence of surviving lineages, their potential origins and proposes a timeframe for their expansions (Karmin et al. 2015; Poznik et al. 2016) (Figure 2). Furthermore, the two haploid non-recombining genetic loci – the Y-chromosome and mtDNA – offer possibility to follow independently patrilineal and matrilineal descents of the extant human (and, using aDNA, also extinct) human populations, impossible to decipher from autosomal genomes. This property makes them unique tools for the reconstruction of our demographic past.

2.4.1. Nomenclature development of Y-chromosomal haplogroups

Initially, there were only a few known binary polymorphisms that could be genotyped by Sanger sequencing or restriction fragment length polymorphism (RFLP) technique. In 1997 a new method (denaturing high-performance liquid chromatography) was used to uncover a large number of novel and informative polymorphisms that could be used in phylogenetic studies (Underhill et al. 1997). However, parallel studies of different scientific groups brought several alternative naming systems that were initially incomparable. First uniform nomenclature was proposed in 2002 by the Y Chromosome Consortium (YCC) (The Y Chromosome Consortium 2002). Nomenclature is constantly renewed and updated with novel markers (Jobling and Tyler-Smith 2003; Karafet et al.

2008; Van Oven et al. 2014). The naming approach proposed by YCC considers monophyletic clades (haplogroup) and assigns capital letters to major clades starting with "A" and moving along the alphabet. Nested sub-clades are subsequently defined by a combination of numbers and lowercase letters (e.g. C3a). Paragroups, that are potentially paraphyletic and derived only on the level of basal nodes are marked with an asterisk, (C3*, called "C3 star"). Such unresolved paragroups may contain sub-lineages and clades, that will become apparent on a deeper resolution level.

Besides YCC, International Society of Genetic Genealogy (ISOGG) (https://isogg.org/tree/) is constantly revising the nomenclature by adding new markers. Because of the rapid accumulation of markers, haplogroup names, defined as amalgamations of letters and numbers, have become unnecessarily long and impractical, e.g. C2a1a1a1a1a (ISOGG Y-DNA tree 2019–2020, https://isogg.org/tree/). This problem can be overcome with another naming strategy offered by YCC (The Y Chromosome Consortium 2002), which maintains the major haplogroup information, but sub-clades are defined by the label of the last defining mutation (marker) e.g. C3-M217 (The Y Chromosome Consortium 2002). In the light of SRS and the great number of newly discovered mutations, this naming convention has become practical to implement.

Current thesis adheres, where possible, to nomenclature suggested by Karmin et al. (Karmin et al. 2015). This system keeps the main principles set by YCC (The Y Chromosome Consortium 2002), combines these with rules proposed by van Oven et al. (Van Oven et al. 2014) and further simplifies the chrY haplogroup nomenclature by considering that alphanumeric length is roughly proportional to the time depth of corresponding haplogroups (Karmin et al. 2015).

2.4.2. Global phylogeny of Y-chromosomal haplogroups

Global phylogeny of chrY (Figure 2) contains 20 main hgs and several sub-clades with oldest describing African variation followed by younger lineages characterising variation outside of Africa (Hallast et al. 2014; Karafet et al. 2008; Karmin et al. 2015; Van Oven et al. 2014; Poznik et al. 2016).

The exclusive African origin of the deepest lineages (A00, A0, A1–3, B) (Barbieri et al. 2016; Batini et al. 2011; Mendez et al. 2013; Scozzari et al. 2014; Wood et al. 2005) in the phylogenetic tree supports the Out of Africa model (Cruciani et al. 2002; Cruciani, Trombetta, Antonelli, et al. 2011; Hallast et al. 2014; Poznik et al. 2016; Stringer 2002; Underhill et al. 2001). According to this model anatomically modern humans arose in Africa and a sub-group of them emerged from the continent settling the rest of the world.

Sister-clade to hg B is CT-M168 that includes most contemporary hgs and splits into two sub-clades – a smaller DE-M145 and a larger CF-P143 – around 75–76 kya (Figure 2). Recently, a novel lineage D0 was described in 3 Nigerian

men (Haber et al. 2019) who were formerly designated as DE* (Weale et al. 2003). This finding also validates recent African origin because D0 is an outgroup to all other hg D chromosomes. Although DE sub-group E-M96 that bifurcates to a broader E1 and minor E2 sub-clades is generally African, it has in time spread out to Middle Eastern and European populations (Battaglia et al. 2009; Cruciani et al. 2004, 2007; Francalacci et al. 2013; King et al. 2011; Semino et al. 2004). While being a sister-clade of E, the very rare D0 is African, the main sub-clade D-M174 has a fragmented spread in East Asia, being constrained to a few populations (Shi et al. 2008; Thangaraj et al. 2005; Watanabe et al. 2019).

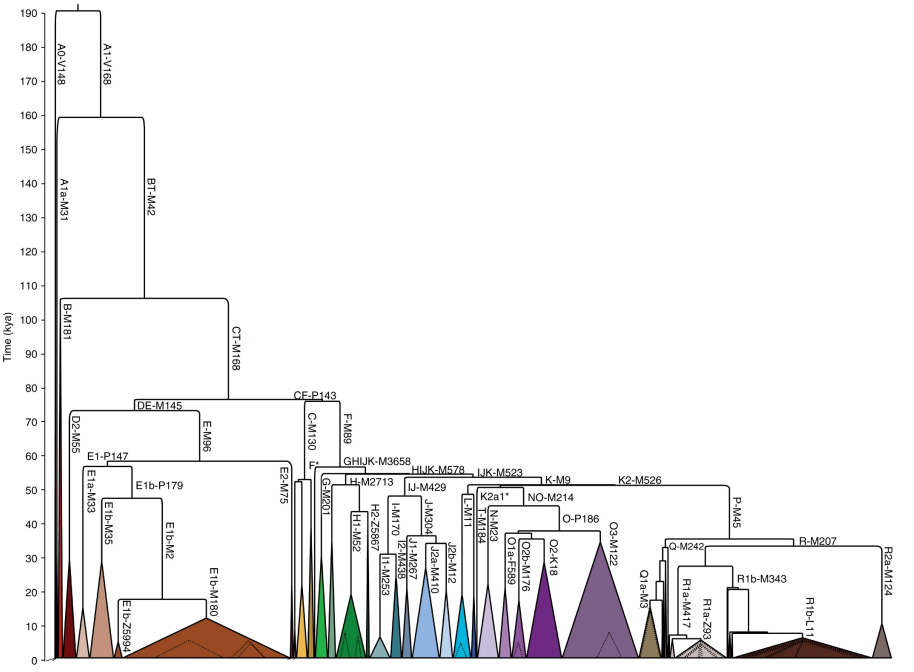

Figure 2. Calibrated global Y chromosome phylogeny.
The phylogenetic tree is based on 1,244 human chrY sequences from 1000 Genomes Project and more than 65,000 variants. Coloured triangles signify major clades and the width corresponds to the sample size. The tree is calibrated using mutation rate $7.6x10^{-10}$ mut/base/year (Fu et al. 2014). Branch lengths are proportional to the time estimations and triangle heights represent coalescence times. Haplogroup names and defining markers are indicated on the branches. Modified from Figure 2 (Poznik et al. 2016) with permission from Springer Nature.

Hg CF-P143 encompasses the non-African variation (Figure 2) (Karmin et al. 2015; Poznik et al. 2016). This includes a mega-haplogroup that splits to F* and GHIJK-M3658. Last of which includes sub-clades G-M201 and HIJK-M578. Clade HIJK-M578 contains hgs H-M2713 and IJK-M523 with two

sub-groups: IJ-M429 and a large K-M9. Hg K-M9 comprises of sub-clades P1-M45 and NO1-M214 (Poznik et al. 2016). Hg P1-M45 sub-clades Q-M242 and R-M207 split around 30–35 kya compared to ~40–44 kya for N-M23 and O-P186. Although hgs N and O are sister-clades, so far, the oldest detected split within hg O predates the oldest detected within hg N by ~10 ky. Hg IJ-M429 diverges into sub-clades I-M170 and J-M304 *ca*. 40–47 kya (Poznik et al. 2016). Haplogroups C-M130, IJ-M429, NO1-M214 and P1-M45 are most important subclades in the Eurasian context.

2.4.3. Phylogeny and phylogeography of haplogroup N and its sub-clades

Paternal haplogroup N-M231 has been studied for more than 20 years. First hg N marker, TAT (M46), was described and genotyped at the end of 1990's (Zerjal et al. 1997). From that time on the number of hg N defining markers and the corresponding phylogenetic resolution has increased. Phylogenetic and phylogeographic studies of hg N have involved both wide geographical areas (Derenko et al. 2007; Pakendorf et al. 2006; Rootsi et al. 2007) and smaller regions (Lappalainen et al. 2006; Tambets et al. 2004). Later, chrY full sequences have been used to update the inner-structure of hg N phylogeny and date the splits within it (Karmin et al. 2015; Poznik et al. 2013, 2016).

Haplogroup N is geographically widespread and covers most of North Eurasia (Figure 3) encompassing many linguistically and culturally different populations (Balanovsky et al. 2008; Chiaroni et al. 2010; Derenko et al. 2007; Karafet et al. 2002, 2018; Karlsson et al. 2006; Karmin et al. 2015; Lappalainen et al. 2006, 2008; Mirabal et al. 2009; Pakendorf et al. 2006; Rootsi et al. 2007; Tambets et al. 2004). Based on some aDNA evidence, it has been speculated that hg N arose in China as it was prevalent in its northeast region in the Neolithic period (Cui et al. 2013; Li et al. 2011; Shi et al. 2013). However, a recent aDNA study (Ning et al. 2020) about Neolithic and Post-Neolithic northern China did not reveal any hg N in a sample set of 27 Y-chromosomes. Therefore, the suggestion that hg N arose during the Neolithic in China is hardly conclusive because documented initial splits within hg N phylogeny occurred long before, around the Last Glacial Maximum (Karmin et al. 2015). In fact, one must admit that both the NO split into hgs N and O some 40 kya (Karmin et al. 2015), as well as early splits within hg N, remain obscure as a far as their phylogeography is concerned, though the "counter-clockwise hypothesis" (Rootsi et al. 2007), is still a possibility. One may also notice that nowadays hg N is rare in China (Shi et al. 2013; Zhong et al. 2011).

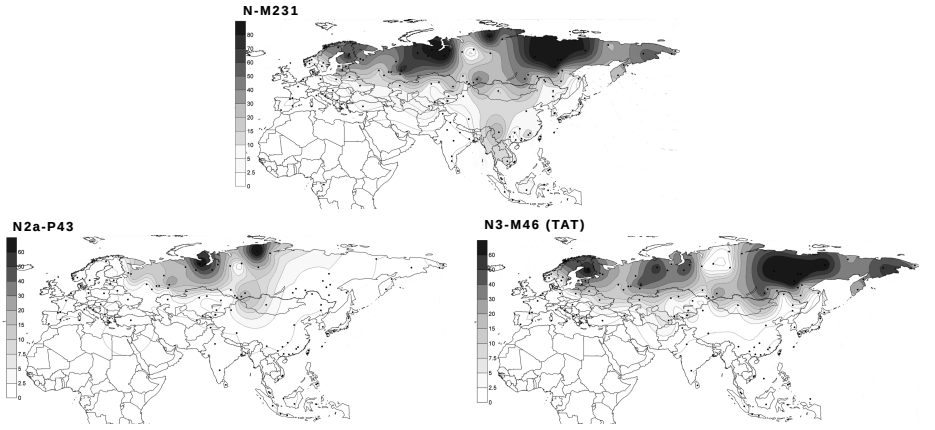

Figure 3. Spatial frequency maps of haplogroup N and its sub-clades N2-P43 and N3-M46.
Dots represent geographic locations of the samples used in the construction of the maps. Frequencies used in the construction of the maps can be found in the supplementary of the original publication. Modified from Figure 2 (Rootsi et al. 2007) with permission from Springer Nature.

The largest sub-clade (Figure 4) of hg N-M231 is N1'2'3-L729 that splits into N1'2-L666 and N3-M46 (TAT) (Karmin et al. 2015). Sub-clade N-M128 is very rare and can be found in minimal frequencies in Japan, Vietnam and Korea (Hammer et al. 2006; Inagaki et al. 2002; Rootsi et al. 2007). The sister-clade of N-M128 is N-P43 which has a wider spread (Figure 3) and has been reported as the main hg around Lake Baikal in South Siberia during the Neolithic period (de Barros Damgaard et al. 2018). Hg N-P43 is common in populations inhabiting the northern parts of Siberia and reaches frequencies of ~90% among Nganasans, up to 75% in Tundra Nenets and around 40% in Forest Nenets (Rootsi et al. 2007). However, ethnically quite similar Selkups possess a frequency of only about 7% (Rootsi et al. 2007). Several other Uralic populations inhabiting the Volga-Ural region (VUR) and areas around southern Urals display a substantial frequency of this clade (Rootsi et al. 2007). In addition, hg N-P43 is common in Turkic populations like Tuvans, Chuvashis and Khakas, and Tungusic-speaking Evens and Evenks (Derenko et al. 2007; Rootsi et al. 2007).

Sub-clade N-P43 is practically missing from populations who inhabit north-eastern parts of Europe and Russian Far East where hg N-M46 (TAT) is predominant (Figure 3). Haplogroup N-M46 (TAT) has a much broader geographical expanse compared to other hg N sub-clades (Balanovsky et al. 2008; Derenko et al. 2007; Fedorova et al. 2013; Karafet et al. 2018; Karlsson et al. 2006; Karmin et al. 2015; Lappalainen et al. 2006; Rootsi et al. 2007; Tambets et al. 2004, 2018) and is frequent among populations inhabiting north-eastern Europe (Finns, Saamis, Estonians, Latvians) (Hallast et al. 2014; Karmin et al. 2015; Rootsi et al. 2007; Tambets et al. 2004), Volga Ural region

(Maris, Mordvin, Komis, Udmurts, Chuvashes, Tatars) (Derenko et al. 2007; Karmin et al. 2015; Rootsi et al. 2007) and Siberia (Nenets, Dolgans) (Karafet et al. 2018; Rootsi et al. 2007). Hg N-M46 sub-clades are also common among Turkic speaking Yakuts with a frequency of over 90%, evidence of a strong recent bottleneck event (Derenko et al. 2007; Fedorova et al. 2013; Karafet et al. 2018; Pakendorf et al. 2006; Rootsi et al. 2007) and belong to a specific clade defined by marker M1979 (Karmin et al. 2015). Hg N-M46 can also be found among populations residing in Siberia (e.g. Buryats, Chukchis, Koryaks) (Derenko et al. 2007; Karmin et al. 2015; Rootsi et al. 2007) and Mongolia (Derenko et al. 2007; Karmin et al. 2015; Rootsi et al. 2007).

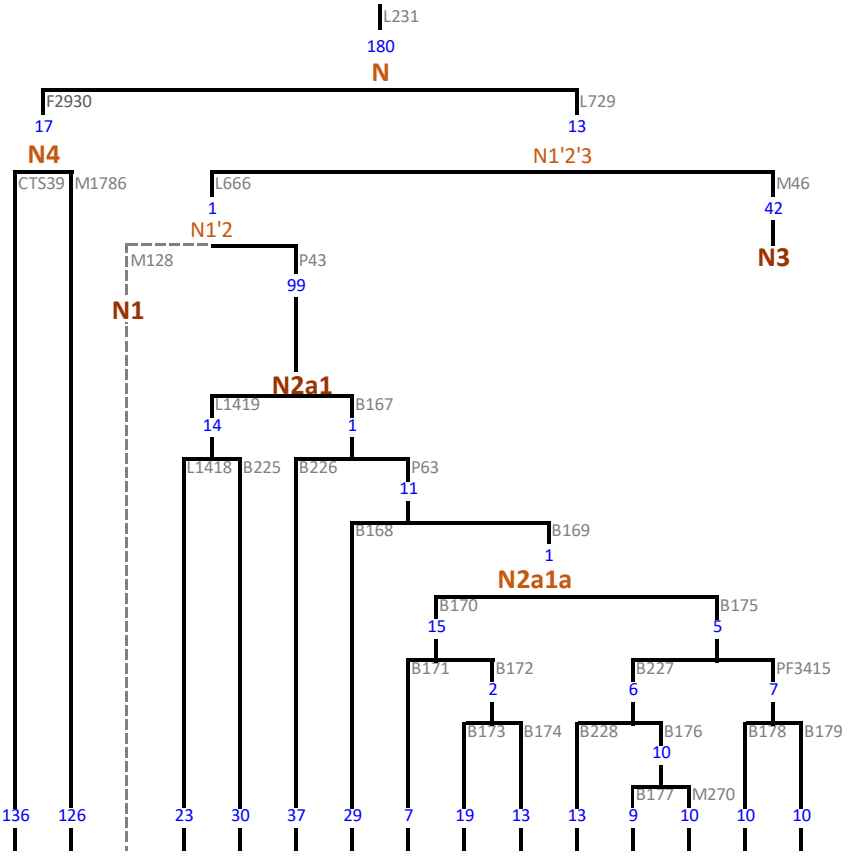

Figure 4. Hg N1'2 cut-out from refined topology of haplogroup N.
Sub-clade names in brown and defining markers in grey are indicated. Mutation counts are shown in blue. Grey dashed lines show known branches with no sequences available in the original publication or low coverage data. Modified from Figure S29 (Karmin et al. 2015). Licensed under the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), © 2015 Karmin et al.

2.4.4. Phylogeography of haplogroup C

Paternal haplogroup C-M130 is one of the main and most extensive hgs in Asia with a wide distribution area extending from the eastern to the central parts of the continent and reaching also Europe (Figure 5) (Bergström et al. 2016; Chiaroni et al. 2010; Gayden et al. 2007; Hammer et al. 2006; Huang, Wei, et al. 2018; Hudjashov et al. 2018; Karmin et al. 2015; Katoh et al. 2005; Kayser et al. 2003, 2006; Lell et al. 2002; Malyarchuk et al. 2010, 2016; Poznik et al. 2016; Sengupta et al. 2006; Wei, Yan, et al. 2018; Xue et al. 2006; Zhong et al. 2010, 2011).

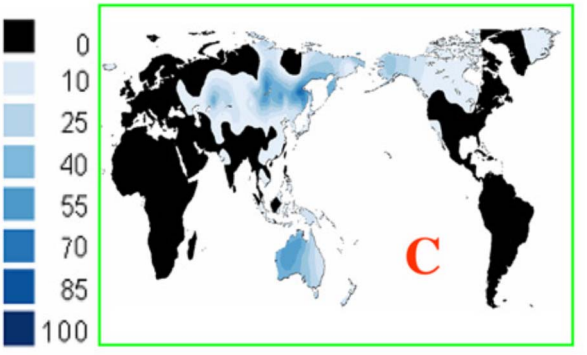

Figure 5. Frequency distribution map of Y-chromosomal haplogroup C.
Modified from Figure 2 (Chiaroni et al. 2010) with permission from PNAS.

Haplogroup C-M130 has two big clusters C1'7-K29 and C3-M217 (naming according to Karmin et al. 2015) (Figure 6) that divided around 50–52 kya (Karmin et al. 2015; Poznik et al. 2016). Sub-lineage C-M8 is scarce and appears to be only present in Japan (Hammer et al. 2006; Karmin et al. 2015; Malyarchuk et al. 2010; Poznik et al. 2016; Zhong et al. 2010). Compared to C-M8 its sister-clade C2'7-K30 (Karmin et al. 2015) has a wide spread with sub-clade C-M38 covering Indonesia, New Guinea, Melanesia and Polynesia (Hudjashov et al. 2018; Karmin et al. 2015; Kayser et al. 2003, 2006; Malyarchuk et al. 2010; Zhong et al. 2010). Sub-lineage C-M347 is exclusive to Aboriginal Australian males with frequency of >40% (Bergström et al. 2016), while clade C-M356 is scarcely spread in South Asian populations (Gayden et al. 2007; Sengupta et al. 2006; Zhong et al. 2010). Recently two new hg C sub-clades C7-B65 and C9-B68 were reported in populations from Maritime Southeast Asia (Karmin et al. 2015).

Haplogroup C3-M217 is the largest C-M130 sub-clade that has spread to many populations residing in a geographically extensive region from North Asia to the Americas (Balaresque et al. 2015; Hammer et al. 2006; Karafet et al. 2002; Katoh et al. 2005; Lell et al. 2002; Malyarchuk et al. 2010; Sengupta et al. 2006; Xue et al. 2006; Zhong et al. 2010). Hg C3-M217 is a significant

haplogroup in South Siberian and Far East populations (Karafet et al. 2002; Lell et al. 2002). For example, sub-clade C3-M407 is quite abundant among Mongolic speaking Buryats (Huang, Wei, et al. 2018; Malyarchuk et al. 2010, 2016), and C3-M48 (sub-clade C3-M77) can be found in Tungusic-speaking Evens and Evenks (Karmin et al. 2015; Lell et al. 2002; Malyarchuk et al. 2010) inhabiting northern and eastern parts of Siberia and Russian Far East.

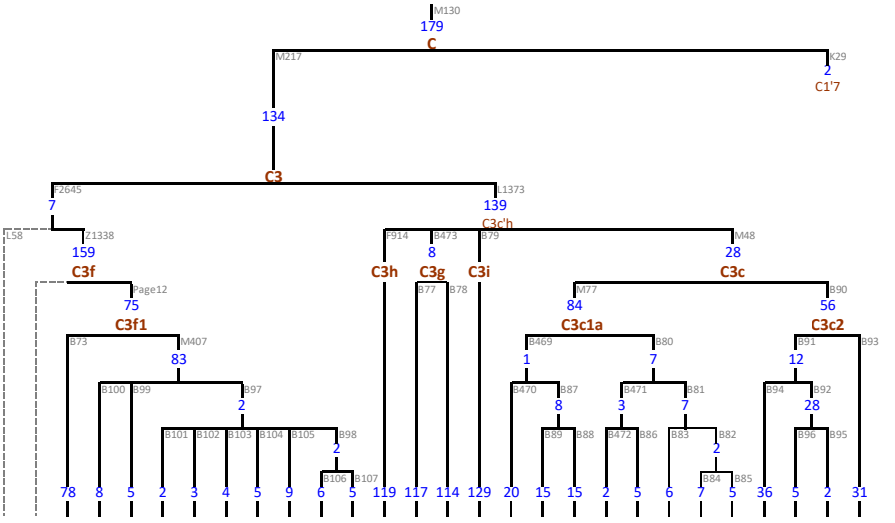

Figure 6. Hg C3 cut-out of refined topology of haplogroup C.
Sub-clade names in brown and defining markers in grey are indicated. Mutation counts are shown in blue. Grey dashed lines show known branches with no sequences available in the original publication or low coverage data. Modified from Figure S20 (Karmin et al. 2015). Licensed under the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), © 2015 Karmin et al.

Hg C3-M217 is also one of the main chrY lineages next to hg O-M175 in East Asia, having the highest frequency in Mongolia averaging around 60% (Huang, Wei, et al. 2018; Katoh et al. 2005; Malyarchuk et al. 2010, 2016; Wei, Yan, et al. 2018; Xue et al. 2006; Yamamoto et al. 2013). Furthermore, aDNA studies of ancient nomadic tribes from the region suggested that C3-L1373 and its sub-clade C3-B78 (F3830, F3889) were important lineages among Donghu nomads (ancestors of contemporary Mongolians) who inhabited parts of Northeast China around 1600–200 BC (Li et al. 2018; Zhang et al. 2018).

Importantly, C3-M217 (name according to Karmin et al. 2015) reaches 8% frequency among populations inhabiting Central Asia (Abilev et al. 2012; Dulik et al. 2011; Malyarchuk et al. 2010; Wei, Yan, et al. 2018; Wells et al. 2001; Zerjal et al. 2003). It is common in Kazakhstan (Wei, Yan, et al. 2018; Wells et al. 2001) especially among the Kereys clan with a frequency of about 76% (Abilev et al. 2012). Compared to Kazakhs from Kazakhstan, the

Kazakhs living in the Altaian Mountains have a lower frequency of hg C3 averaging to ~40% (Dulik et al. 2011; Malyarchuk et al. 2010). Sub-clades of C3-M217 are also modestly present among some European populations (Huang, Wei, et al. 2018; Malyarchuk et al. 2010; Wei, Yan, et al. 2018).

A well-known hypothesis related to hg C3 has been proposed by Zerjal et al. (2003) linking it to Genghis Khan and his male descendants, attributing its spread alongside the Mongol Empire expansions. More precisely, the evidence came from studying STR profiles of hg C3*(xC3c) in numerous Central and Inner Asian populations in a search of star-like, i.e. relatively recent clusters, coalescing within the last millennium. One such cluster (see e.g. Figs 1 and 2 in Zerjal et al. (2003)) that coalesces about 1000 years ago, while aspects of its phylogeography, like abundance not only among Mongols but also Hazara living predominantly in Afghanistan and claiming their patrilineal descent from one of the Genghis Khan "generals", who presumably belonged to the same clan as the famous khan. However, there is no exact evidence to the claim and several studies have expressed caution towards the notion of Genghis Khan (clan) as the sole source of the clade under discussion (Abilev et al. 2012; Batbayar and Sabitov 2012; Dulik et al. 2011; Malyarchuk et al. 2010; Wei, Yan, et al. 2018; Zakharov 2010; Zhang et al. 2018). In more general terms, it awoke an interesting quantitative discussion whether cultural transmission of reproductive success (Genghis Khan and alike, see e.g. (Balaresque et al. 2015)) is the only possible explanation for such phenomena – widely spread recent star-like STR modal haplotype clusters. It was shown (Guillot and Cox 2015) by coalescent simulation that high frequency haplotypes can appear just by chance. In short, though cultural transmission of reproductive success is of course possible, it is not "statistically needed".

2.4.5. Dispersal of other paternal haplogroups in Northern and Central Eurasia

In addition to previously mentioned extensive hgs C and N, there are several other lineages with varying frequencies present among the diverse populations that inhabit northern and central regions of Eurasia. Most widespread hgs include hgs O, R1, I1 and I2, while hgs Q, G and J2 are less common in the mentioned area and exhibit a more localized spread pattern

Haplogroup O-M175, the sister-clade to N-M231, and its sub-cades are most common in Southeast Asia (Cai et al. 2011; Chiaroni et al. 2010; Karmin et al. 2015; Poznik et al. 2016; Su et al. 1999; Wang and Li 2013; Wei et al. 2017; Zhong et al. 2011) and Melanesia in populations from Solomon Islands, Fiji, Moluccas (Maluku Islands) and Papua New Guinea (Chiaroni et al. 2010; Cox and Lahr 2006; Delfin et al. 2012; Kayser et al. 2003, 2006; van Oven et al. 2014). However, it also extends to Mongolia and somewhat to Kazakhstan where frequencies are between 5–10% (Katoh et al. 2005; Wang et al. 2013; Wang and Li 2013; Wells et al. 2001). The phylogeny of hg O is quite complex

with 2 major sub-clades (O1'2-F190 and O3'7-M122) that branch off to several smaller and larger sub-lineages (Karmin et al. 2015; Poznik et al. 2016). While O1-M119, O2-M95 and O3'7-M122 (names according to Karmin et al. 2015) are frequent in Southeast Asia and Polynesia (Cai et al. 2011; Hudjashov et al. 2018; Su et al. 1999; Trejaut et al. 2014; Wei et al. 2017), the only hg O sub-clade present Mainland Southeast Asia Eurasia is O-M122 that is the most common sub-lineage in China (Shi et al. 2005; Wang et al. 2013; Wang and Li 2013; Wells et al. 2001).

Haplogroup Q sub-clade Q1-L330 is quite widespread in several North Eurasian populations (Grugni et al. 2019; Karmin et al. 2015; Pinotti et al. 2019; Wei, Wang, et al. 2018). Hg Q-P36/M242 (Q1-L330, (according to Karmin et al. 2015) is especially frequent in some Siberian ethnic groups like Kets (>95%) and Selkups (>66%) (Karafet et al. 2002). This lineage is also substantial in Mongolian populations (Battaglia et al. 2013; Grugni et al. 2019). Haplogroup Q2-F1096 (according to Karmin et al., 2015) and its sub-lineages (M120, M25, B143) are present in Northeast Siberia, Western Eurasia, Central Asia and Mongolia (Balanovsky et al. 2017; Battaglia et al. 2013; Karafet et al. 2018; Karmin et al. 2015). Furthermore, hg Q can also be found in different populations from the Altai region. Recently, phylogeny of Q3-L275 was resolved with sub-lineages detected in some European (Balanovsky et al. 2017; Grugni et al. 2019; Huang, Pamjav, et al. 2018) as well as North and Central Asian populations (Balanovsky et al. 2017; Huang, Pamjav, et al. 2018). Q-M242 is also a major Native American haplogroup. Sub-lineage Q1a-M3 is almost entirely constrained to the Americas (Battaglia et al. 2013; Grugni et al. 2019; Karafet et al. 2008; Lell et al. 2002; Pinotti et al. 2019; Wei, Wang, et al. 2018; Zegura et al. 2004).

Hg Q sister-clade R-M207 and especially its sub-clade R1-M173 is one of the main and most widespread haplogroups in Eurasia (Balaresque et al. 2010; Battaglia et al. 2009; Busby et al. 2012; Kushniarevich et al. 2015; Myres et al. 2011; Semino et al. 2000; Underhill et al. 2010, 2015). Haplogroup R1 contains two large sub-clades: R1a-M420 and R1b-M343 (names according to Karmin et al. 2015) (Karmin et al. 2015; Poznik et al. 2016). Hg R1a-M420 (naming according to Karmin et al. 2015) is mostly found in Central/Eastern Europe (Battaglia et al. 2009; Kayser et al. 2005; Kushniarevich et al. 2015; Underhill et al. 2010, 2015) with offshoots to Central Asia, South Siberia and Caucasus (Balanovsky et al. 2011; Kushniarevich et al. 2015; Underhill et al. 2010, 2015; Yunusbayev et al. 2012). Two major R1a sub-clades, R1a1-Z283 (Z282, name according to Underhill et al., 2015) and R1a2-Z93, display separate geographical distributions (Figure 7) (Underhill et al. 2015). Frequencies of hg R1a1 sub-clades exceed 30% in Eastern and Central European populations (Kushniarevich et al. 2015; Rębała et al. 2013; Underhill et al. 2010, 2015). These sub-lineages of R1a1 can also be found among VUR populations (Underhill et al. 2015). Hg R1a2-Z93 can generally be found in Central and South Asia as well as South Siberia especially in the Altai region (Underhill et al. 2015). Hg R1b-M269 (R1b-M343 according to Karmin et al.

2015), however, is mostly found in Western and Southern Europe (Balaresque et al. 2010; Batini et al. 2015; Busby et al. 2012; Cruciani, Trombetta, Antonelli, et al. 2011; Gonçalves et al. 2006; Larmuseau et al. 2014; Myres et al. 2011; Zalloua et al. 2008). Occurrence of R1b diminishes towards Eastern Europe and the Balkans (Batini et al. 2015; Myres et al. 2011).

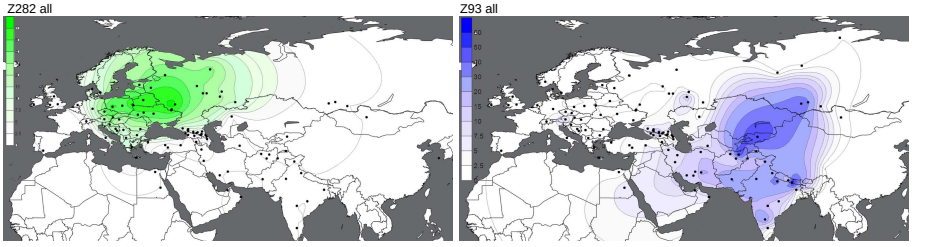

Figure 7. Spatial frequency maps of R1a sub-clades R1a1-Z282 and R1a2-Z93. Dots represent sampling points. Modified from Figures 2 and 3 (Underhill et al. 2015) with permission from Springer Nature.

Another extensive clade in North and Central Eurasia is IJ-M429 that splits to I-M170 and J-M304 (Karmin et al. 2015; Poznik et al. 2016). These sister-clades mostly display separate geographic distributions with some exceptions (Chiaroni et al. 2010; Rootsi et al. 2004; Semino et al. 2004; Underhill et al. 2007). Sub-clade I-M170 is generally constrained to Europe, making up around 18–20% of the paternal lineages in the region (Figure 7) (Alvarez et al. 2014; Barać et al. 2003; Batini et al. 2015; Battaglia et al. 2009; Fechner et al. 2008; Gonçalves et al. 2006; Lappalainen et al. 2006; Passarino et al. 2002; Rootsi et al. 2004; Semino et al. 2000; Underhill et al. 2007). It has two noticeable frequency peaks. One in Northern Europe (Scandinavia) where sub-clade I1-M253 is frequent (Lappalainen et al. 2006; Rootsi et al. 2004; Underhill et al. 2007), and the other in West Balkans (Dinaric Alps) where I2a-M423 is commonplace (Battaglia et al. 2009; Underhill et al. 2007). In some parts of Eastern and Central the geographic distribution areas of I1a and I2a overlap (Underhill et al. 2007).

Haplogroup J-M304 is mostly restricted to Southern Europe, Near East and Northern Africa with off-shoots to South Asia (Al-Zahery et al. 2003, 2011; Chiaroni et al. 2010; Finocchio et al. 2018; Kivisild et al. 2003; Manco et al. 2018; Semino et al. 2004; Singh et al. 2016). In the phylogeny of hg J there are two main sub-clades: J1-M267 and J2-M172 (Karmin et al. 2015; Poznik et al. 2016). Sub-clade J1-M267 is very sporadic in Europe with minor presence mostly in southern parts (Batini et al. 2015; Battaglia et al. 2009; Manco et al. 2018; Semino et al. 2004). Its sister-clade J2-M172 is prevalent within Europe especially in the southern European and Balkan populations (Batini et al. 2015; Battaglia et al. 2009; Di Giacomo et al. 2004; Manco et al. 2018; Semino et al. 2004; Zalloua et al. 2008). In some instances, the frequencies can reach over 20%, however, they drop dramatically towards

Central and Northern Europe (Batini et al. 2015; Battaglia et al. 2009). Other regions where hg J2 is rather abundant include the Caucasus, Turkey and Near East (Balanovsky et al. 2011; Di Giacomo et al. 2004; Hovhannisyan et al. 2014; King et al. 2011; Nasidze et al. 2003, 2009; Semino et al. 2004; Yunusbayev et al. 2012; Zalloua et al. 2008).

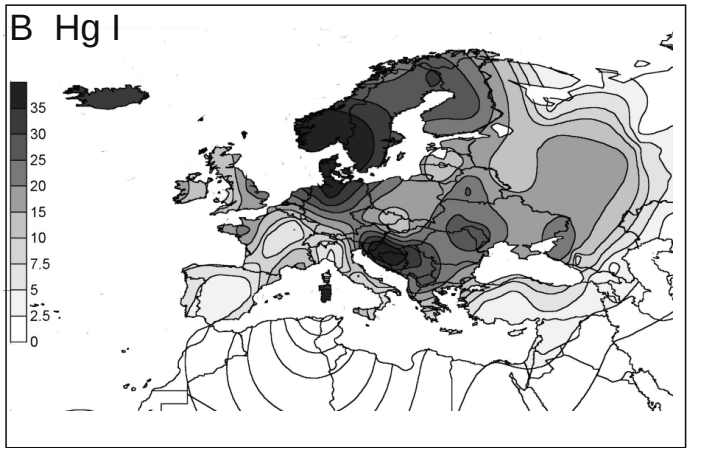

Figure 8. Frequency distribution map of haplogroup I-M170.
Frequencies used in the construction can be found in the original publication. Modified from Figure 1 (Rootsi et al. 2004) with permission from Elsevier.

Hg G-M201 has a constrained distribution covering mostly the Caucasus, Near/ Middle East and Southern Europe (Balanovsky et al. 2011; Battaglia et al. 2009; Capelli et al. 2007; Chiaroni et al. 2010; Cristofaro et al. 2018; Hernández et al. 2019; Rootsi et al. 2012; Voskarides et al. 2016; Yunusbayev et al. 2012). G-M201 splits into two sub-clades: G1-M285 and G2-P287 (Balanovsky et al. 2015; Behar et al. 2017; Hallast et al. 2014; Karmin et al. 2015; Poznik et al. 2016). Sub-branch G1-M285 is scattered over a large region and can be found in higher frequencies among populations of Southwest and Central Asia (Balanovsky et al. 2015). The sister-clade G2-P287 (P15) is more common compared to G1, and most widespread in the Caucasus where it averages around 8–12% (Balanovsky et al. 2011; Rootsi et al. 2012; Yunusbayev et al. 2012). In Europe G2 can be found in southern regions like Corsica and Northern Italy (Battaglia et al. 2009; Rootsi et al. 2012).

2.5. Two exceptional populations in Europe

2.5.1. Hungarians – a Uralic-speaking population in Central Europe

2.5.1.1. General information and history

Europe today is mostly inhabited by people speaking Indo-European languages. However, Hungarians who reside in Central Europe, speak a Uralic language. The geographically nearest members of the same linguistic family live in north-eastern Europe (Finns and Estonians) and closest linguistic relatives, with whom the Hungarians belong to the Ugric branch of the Uralic languages, are West Siberian Khanties and Mansis (Figure 9). With a population size of around 13 million (https://www.ethnologue.com), Hungarians are the largest population natively speaking a Uralic language. For a long period of time the Hungarian language was strongly influenced by several Turkic languages with first contacts thought to have taken place about 1500 years ago (Róna-Tas and Árpád 2011).

Figure 9. Geographic spread of Uralic-speaking populations.
Reprinted from Figure 1A (Tambets et al. 2018). Licensed under the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

How exactly a population in Central Europe has come to speak a language that is so distinct from the neighbouring Indo-European populations is hard to pinpoint. The history of the Carpathian Basin, especially in the Middle Ages, is extremely complex involving multiple migrations from the Eurasian Steppe by several different groups in a short period of time (Csáky et al. 2020; Fóthi

2000; Fóthi et al. 2020; Neparáczki et al. 2019; Róna-Tas 1999; Türk 2012). Around 400–450 AD the Carpathian Basin was mostly under the control of the nomadic Huns who were defeated by an Eastern Germanic group the Gepids (Neparáczki et al. 2019). In the year 568 AD, the Avars established an empire that occupied a large area from the Carpathian Basin to the Pontic-Caspian Steppes (Csáky et al. 2020; Neparáczki et al. 2019). Although by the beginning of the 9th century the Empire ceased to exist (Csáky et al. 2020; Neparáczki et al. 2019), the Avar population still survived and were unified with the subsequent Hungarian state (Neparáczki et al. 2019). First Hungarians, also known as the Conquerors, reportedly arrived in the Carpathian Basin around 860–870 AD as sporadic evidence of it can be found in the area (Boldog et al. 2015; Neparáczki et al. 2019). In 895 AD the Conquerors occupied the Carpathian Basin and by 907 AD established a permanent rule in the entire region (Boldog et al. 2015; Fóthi et al. 2020; Türk 2012).

The Uralic languages cover an extensive area of North Eurasia. Generally, it is accepted that the homeland of the Uralic linguistic family was either the Volga-Ural region (Koivulehto 2001), or further in Siberia (Janhunen 2009). The Hungarian Conquerors are thought to be the putative ancestors of contemporary Hungarians. It has been suggested that the conquerors included people from several tribes and some of them were of Ugric descent (Fóthi et al. 2020). However, linguistic affiliations and genetics may differ. Ideed, there are many different opinions about the possible territories from where the ancient Hungarians may have originated (see, e.g., Fórhi et al 2020). One of them being West Siberia/Southern Urals. Among the others are Altai Mountains and Black-Sea/Northern Caucasus. Contemporary archaeological finds from the southern Ural region and around Dnieper river in Ukraine indicate that the migration of the Conquerors was fast (Boldog et al. 2015). In addition, the sites of the Chiyalikskaya culture contain archaeological finds that point to Hungarian ancestors living in the Ural region up to the 13th century (Fodor 1977; Garustovic 1988; Kazakov 1986; Klíma 2016; Türk et al. 2015). First universally accepted cultures, located to the west of the Urals and connected with the early Hungarian migration route, are Kushnarenkovo (6th–8th centuries CE) (Fodor 2015; Ivanov et al. 2009; Kazakov 1981; Sedov 1987) and Karayakupovo cultures (9th–10th centuries CE) (Akbulatov 1988; Fodor 2015; Matveeva 2007). Archaeological sites connected with these cultures can mainly be found in present-day Bashkortostan, Tatarstan and the Chelyabinsk area of the Trans-Ural region (Boldog et al. 2015; Türk 2012). Furthermore, early chronicles mention the easterly homeland of Hungarians called *Magna Hungaria* (Great/Ancient Hungary) (Anonymus 2010; Keža 1999).

2.5.1.2. Genetics and paternal gene pool of contemporary Hungarians

The gene pool of present-day Hungarians is almost indistinguishable from their Indo-European speaking geographical neighbours. The autosomal genetic diversity in all Uralic-speaking populations is mostly influenced by geographic proximity and in contrast to most other Uralic speakers, Hungarians do not share their common ancestral Siberian autosomal component (Tambets et al. 2018). In addition, geographically distant Uralic linguistic relatives share more Identical by Descent (IBD) segments with each other than with equally distant populations speaking other, non-Uralic, languages. However, Hungarians do not have excess IBD sharing with Uralic speakers from VUR or Siberia. Furthermore, they form genetic clusters with their non-Uralic geographic neighbours instead of linguistic relatives (Tambets et al. 2018). This means that the recent genetic ancestry of Hungarians has strongly been influenced by their present-day neighbouring populations.

Y-chromosomal haplogroups

Similarly to autosomal (Tambets et al. 2018) and mtDNA data (Egyed et al. 2007), the composition of paternal gene pool of the contemporary Hungarians is similar to neighbouring Central European populations with most common hgs being R1-M173, I-M170, E2-M215 and J2-M172 (names according to Karmin et al., 2015) (Bíró et al. 2015; Csányi et al. 2008; Pamjav et al. 2017; Völgyi et al. 2009)

Haplogroup R1 sub-clades R1a-M420 and R1b-M343 (hg names according to Karmin et al. 2015) are both abundant. However, hg R1a is somewhat more common with a frequency over 20% while hg R1b mostly remains ~15–18% (Batini et al. 2015; Battaglia et al. 2009; Bíró et al. 2015; Csányi et al. 2008; Myres et al. 2011; Pamjav et al. 2012; Underhill et al. 2015; Völgyi et al. 2009).

Hungarian hg I-M170 chrYs belong to both I1-M253 (I1a-Z58 in Karmin et al. 2015) and I2a-M423 sub-clades. On account of geographic closeness to the Balkans, I2a is more abundant than I1 (5–8%) with a frequency ~16–19% (Batini et al. 2015; Battaglia et al. 2009; Bíró et al. 2015; Csányi et al. 2008; Rootsi et al. 2004; Underhill and Kivisild 2007; Völgyi et al. 2009).

As previously mentioned, haplogroup E has expanded to Southern and Central European populations including Hungarians. The most common hg E2-M215 sub-clade is E2a1-V13 (name according to Karmin et al., 2015) with a frequency ~10% (Bíró et al. 2015; Csányi et al. 2008; Pamjav et al. 2017; Völgyi et al. 2009).

Haplogroup J-M304 is also quite frequent in the paternal gene pool of Hungarians (Bíró et al. 2015; Csányi et al. 2008; Pamjav et al. 2017; Völgyi et al. 2009). Although both J1-M267 and J2-M172 have been detected, hg J2 is more abundant with a frequency of ~5–6% (Bíró et al. 2015; Csányi et al.

2008; Völgyi et al. 2009). This can be explained with the proximity to the Balkans where hg J2 is widespread (Batini et al. 2015; Battaglia et al. 2009; Semino et al. 2004; Zalloua et al. 2008).

Besides the major haplogroups that comprise ~80% of the Hungarian paternal gene pool, several rare hgs have been detected. These haplogroups are H1a-M52 (M82), G2a-P15,N3-M46 and Q-M242 (names according to Karmin et al., 2015) (Bíró et al. 2015; Csányi et al. 2008; Pamjav et al. 2017; Völgyi et al. 2009). Haplogroup N3-M46, that is extremely common among Uralic speakers, is rare (~1–2%) among Hungarians (Bíró et al. 2015; Csányi et al. 2008; Pamjav et al. 2017; Rootsi et al. 2007; Völgyi et al. 2009). In one sub-population (Bodrogköz) sub-clades N3a3-VL29 and N3a4-Z1936 (names according to Karmin et al. 2015) have frequencies of ~4% and ~1%, respectively (Pamjav et al. 2017). The presence of hg N3a3 can be connected to geography as Carpathian Basin is in the southern spread border of the sub-clade. However, the same is not true for hg N3a4, that is common among populations in the VUR, Siberia and Northeast Europe (Derenko et al. 2007; Rootsi et al. 2007; Tambets et al. 2004) with a sharp drop in frequency towards southern populations.

Around 2–5% of Hungarians carry the same hg H sub-clade, H1-M82 (Pamjav et al. 2017; Völgyi et al. 2009), as European Roma (Gusmão et al. 2008; Klarić et al. 2009; Sengupta et al. 2006). Haplogroup G2, that is common in the Caucasus (Balanovsky et al. 2011; Yunusbayev et al. 2012) and also present around the Mediterranean (Battaglia et al. 2009; Cristofaro et al. 2018; Hernández et al. 2019; Karafet et al. 2005; Rootsi et al. 2012; Voskarides et al. 2016), has a marginal occurrence in the Hungarian gene pool represented by the sub-clade G2a-P15 (name according to Karmin et al., 2015) (Bíró et al. 2015; Csányi et al. 2008; Pamjav et al. 2017; Völgyi et al. 2009). Compared to other minor hgs, the spread of hg Q-M242, specific to some populations in Western Siberia and the Altai region (Dulik et al. 2012; Huang, Pamjav, et al. 2018; Karafet et al. 2002; Malyarchuk et al. 2011), is patchy with a frequency below 1% (Bíró et al. 2015; Pamjav et al. 2017)

2.5.2. Kalmyks – the only Mongolic-speaking population of south-eastern European Plain

2.5.2.1. General information and history

Kalmyks are the only geographically European population who speaks a Mongolic language, more specifically an Oirat dialect (language) and practices Buddhism. Linguistically Kalmyks are in the Mongolic sub-branch of Altaic languages with Mongols and Buryats (Galushkin et al. 2001; Janhunen 2006; Katzner 1986). Kalmyks inhabit the south-eastern part of the European Plain in Russia, residing in the steppes of the Caspian-Lower Volga River basin (Galushkin et al. 2001; Malyarchuk et al. 2013; Nasidze et al. 2005; Roewer et al. 2007). Size of the population in the Russian Federation varies according

different sources from 160,000 to 180,000 ethnic speakers (https://www.ethnologue.com, https://www.gks.ru). Historical evidence state that the predecessors of Kalmyks were Oirat-speaking groups. These tribes led a nomadic lifestyle and migrated from western parts of Mongolia to Eastern Europe around 400 years ago where they have been in virtual political, geographic and also genetic isolation ever since (Ashilova 1976; Birtalan 2012; Bougdaeva and Isaacs 2018; Galushkin et al. 2001; Khomyakova and Balinova 2017; Malyarchuk et al. 2013; Nasidze et al. 2005; Roewer et al. 2007).

In the Mongol Empire (13th–14th centuries) Oirat rulers often married into the Genghis Khan dynasty, which enabled them to retain the tribal structure (Juvaini 1997). In the 13th century the Durben-Oirat alliance was founded between four Oirat tribes (Torgut, Derbet, Khoshut and Khoit) (Galushkin et al. 2001; Malyarchuk et al. 2013). In the 17th century Torguts and Derbets migrated to the banks of the Volga River in south-eastern European Plain (Galushkin et al. 2001; Malyarchuk et al. 2013; Roewer et al. 2007). The main reason for the relocation was thought to have been the discontent with the centralisation of power and shortage of pasture lands in their homeland (Galushkin et al. 2001; Nasidze et al. 2005). Later these groups founded the Kalmyk Khanate in the lower Volga and became the ethnic group Kalmyks (Galushkin et al. 2001; Nasidze et al. 2005; Roewer et al. 2007). The Kalmyks maintained their nomadic lifestyle for almost 200 years but accepted agriculture by the end of 19th century (Galushkin et al. 2001; Roewer et al. 2007).

Nowadays there are several indigenous groups who speak Oirat dialects and inhabit a wide area of Western Mongolia and Xinjiang Uygur Autonomous Region in China. These groups include Derbet, Torgut, Khoshut, Olot, Dzungar, Bayad, Zakhchin, Khoton, Myangad and Buzava. In the Russian Federation ethnic Oirat speakers comprise of Torguts, Derbets, Buzavas and a small group of Khoshuts who have retained their ethnic tribal names (Galushkin et al. 2001). One population also of Oirat origin is the Kalmaks who live in small groups along the migration routes of the Oirats. The largest group of the Kalmaks is Sart-Kalmaks who reside in Kyrgystan (Balinova 2015).

2.5.2.2. Y-chromosomal gene pool of the Kalmyk population

Genetic studies including chrY research involving the Kalmyk population have been rather limited, hindered by low resolution and small sample sizes (Derenko et al. 2006; Dulik et al. 2012; Galushkin et al. 2001; Malyarchuk et al. 2013; Nasidze et al. 2005; Roewer et al. 2007). However, it has been shown that the chrY haplogroup variation between different Kalmyk ethnic groups is well visible (Malyarchuk et al. 2013).

The predominant hg in the Kalmyk population is C3-M217 which, along with its sub-clades, makes up well over 50% of their paternal gene pool (Malyarchuk et al. 2013; Nasidze et al. 2005). The high frequency connects Kalmyks with different sub-groups of Mongols from Mongolia and Mongolic Buryats of South Siberia (Katoh et al. 2005; Malyarchuk et al. 2010, 2016;

Yamamoto et al. 2013; Zhong et al. 2010). Ethnic Kalmyk sub-groups such as Torguts, Derbets, Khoshuts and Buzawas have high frequency (~40%) of sub-clade C3c1-M77 (Malyarchuk et al. 2013). Another hg C3 subclade – C3f1-M407 (according to Karmin et al. 2015) is significantly less abundant and present among Derbets, Torguts and Buzawas (Malyarchuk et al. 2013). Both sub-clades occur among Mongols and Buryats with hg C3f1-M407 also present in Evens and Evenks (Karmin et al. 2015; Malyarchuk et al. 2010; Zhong et al. 2010).

The other hgs among different Kalmyk groups are N3-M46, O2-P31, O3-M122 and R2a-M124 (names according to Karmin et al. 2015). These constitute ~30% of their paternal gene pool. Intriguingly, hg N1c-TAT (N3-M46, according to Karmin et al. has a significant frequency (almost 40%) among Khoshuts while in other groups the occurrence is considerably lower (~5% in Derbets, ~2% in Torguts) (Malyarchuk et al. 2013). Hg N3-M46 is abundant among north-eastern European, VUR as well as several Siberian populations including Buryats, and can also be found in Mongolian population (Derenko et al. 2007; Rootsi et al. 2007).

Similar is the case with hg O3-M122 that among the Khoshuts reaches a frequency of about 17% but among the Derbets and Torguts stays below 6% (Malyarchuk et al. 2013). Haplogroup O3 is the main hg in East Asia but its frequency in Mongolia is considerably lower than in other Asian populations (Katoh et al. 2005; Su et al. 1999; Wang and Li 2013; Wei et al. 2017; Yan et al. 2011, 2014). Another hg O-M175 sub-clade O2-P31 is copious only among the Torguts with a frequency of ~13% (Malyarchuk et al. 2013). South Asian hg R2-M124 (Kivisild et al. 2003; Sengupta et al. 2006) is quite plentiful (~15%) in Derbets with low frequencies (<1%) in Khoshuts and Torguts (Malyarchuk et al. 2013).

Marginal haplogroups, that have been detected in Derbets and Torguts, include hgs D-M174, J1-M267 and J2a-M410 (Malyarchuk et al. 2013). Different J hgs may be present in their gene pool due to the geographic spread pattern as J can be found in the Caucasus and Near East where this hg is abundant (Al-Zahery et al. 2011; Balanovsky et al. 2011; Nasidze et al. 2003; Semino et al. 2004; Yunusbayev et al. 2012). Additionally, several hg Q sub-clades like Q1a1-M120, Q1a2-M25 and Q1a3-M346 (Q2c-M120, Q2a-M25, Q1-M346 according to Karmin et al. 2015) are also present in the ethnic sub-groups of Kalmyks. Hg Q1a3 is similarly to hgs D, J1 and J2 found among the Derbets and Torguts. Hg Q1a2-M25 occurs only in Khoshuts and Q1a1-M120 only in Derbets (Malyarchuk et al. 2013). These Q sub-lineages may have come to the Kalmyk gene pool from the Altai region or South Siberia where this hg is common and where Mongolic ethnic groups have lived throughout history (Battaglia et al. 2013; Dulik et al. 2012; Malyarchuk et al. 2011). Furthermore, hg N1b-P43 (N2a1-P43, according to Karmin et al. 2015) has a low frequency (<2%) among all the Kalmyk sub-groups (Malyarchuk et al. 2013). Hg R1-M173 and its subclades exhibit a similar pattern of minor frequencies below 4% in the Kalmyk groups (Malyarchuk et al. 2013).

2.6. Ancient Y-DNA research

The DNA of historic and pre-historic individuals or ancient DNA (aDNA) provides the opportunity to garner genetic information about the specific periods and regions where the samples were collected from. The development of Next Generation Sequencing technologies (Margulies et al. 2005) (NGS) has transformed aDNA research which now uses whole genomes of ancient individuals to characterise genetic variation. NGS allows the sequencing of short DNA fragments (Margulies et al. 2005) which is beneficial in the case of aDNA as it degrades over time (Orlando et al. 2015; Pääbo 1989). Radiocarbon dating of ancient samples adds a specific time point. Along with archaeological background data, it is possible to describe the demographic processes that affect the genetic structure of populations and the formation of their gene pool on a temporal scale.

Although aDNA is a very beneficial tool in population genetics, it has quite a few challenges such as post-mortem damage, low content of endogenous DNA, and the limited amount of material available. Furthermore, due to the degraded nature of aDNA, it is important to avoid modern day contamination by working in clean-labs and wearing protective suits (Knapp et al. 2012; Orlando et al. 2015).

Ancient Y-DNA studies have improved our knowledge of the past paternal variation. Although sequencing ancient chrY and getting adequate coverage can be challenging, MSY sequences of ancient individuals have provided insight into the history of chrY variation. For example, ancient Y-DNA sequences of dated human remains can be used to see if they are consistent with conclusions made from phylogenetic trees built using modern data. Additionally, ancient Y-DNA may also shed light on clades that have gone extinct and cannot be detected from contemporary chrY sequences (Kivisild 2017).

2.6.1. Overview of ancient Y chromosome studies in Europe

The formation of European paternal gene pool has been extensively characterised by many ancient DNA studies (Allentoft et al. 2015; Brandt et al. 2015; Fernandes et al. 2018; Fu et al. 2015, 2016; Gamba et al. 2014; González-Fortes et al. 2017; Günther et al. 2015, 2018; Haak et al. 2015, 2010; Hofmanová et al. 2016; Järve et al. 2019; Jones et al. 2015; Keller et al. 2012; Krzewińska et al. 2018; Lacan et al. 2011; Lamnidis et al. 2018; Lazaridis et al. 2014, 2016; Lee et al. 2012; Lipson et al. 2017; Martiniano et al. 2017; Mathieson et al. 2015, 2018; Mittnik et al. 2018; Neparáczki et al. 2017, 2019; Olalde et al. 2014, 2018; Omrak et al. 2016; Pereira et al. 2017; Saag et al. 2017, 2019; Sánchez-Quinto et al. 2019; Seguin-Orlando et al. 2014; Sikora et al. 2014, 2017; Szécsényi-Nagy et al. 2015).

Anatomically modern humans populated Europe around 45 kya (Benazzi et al. 2011; Kivisild 2017; Lazaridis 2018) and spread over the whole continent during the Upper Palaeolithic (Fu et al. 2016; Kivisild 2017; Lazaridis 2018). Some of the earliest samples dated to 30–40 kya were found from several sites in Romania (Fu et al. 2015, 2016; Trinkaus et al. 2003). Two male individuals from these locations belonged to hgs F (Fu et al. 2015) and DT (Fu et al. 2016). Additionally, ancient individuals were uncovered from burial sites on the northern slope of the Pavlovské Hills in Czech Republic dated to 30–31 kya (Fu et al. 2016). One of these males was derived for hg C1a2 and other two were assigned to branches leading to hgs CT and BT (Fu et al. 2016).

During the Mesolithic, hunter-gatherers (HG) were dominant in Europe. A 14 ky-old male from one of the earliest sites in Italy was assigned to hg R1b (Fu et al. 2016). Mesolithic HG males from the Iron Gates gorge in Romania and Serbia were dated to ~9–11 kya and carried chrY hgs R and I with most prominent sub-lineages being R1b and I2a (González-Fortes et al. 2017; Mathieson et al. 2018). Individuals from German sites (~15–16 ky) were also derived from different I-M170 sub-clades (Fu et al. 2016; Mathieson et al. 2018). Hunter-gatherers from the Baltics carried R1a and I haplogroups (Mittnik et al. 2018; Saag et al. 2017). One HG individual from Estonia belonged to sub-lineage R1a5 (Saag et al. 2017) – a deep R1a subclade that is currently extremely rare (R-YP1276 according to YFull YTree v8.01.00, https://www.yfull.com/). Additionally, two Latvian HG were assigned to hg R1b sub-clades (Jones et al. 2017).

The Neolithic era brought to Europe a population turnover as HG were largely replaced by agriculturalists from Anatolia (Behar et al. 2017; Bollongino et al. 2013; Bramanti et al. 2009; Fernandes et al. 2018; Gamba et al. 2014; Günther et al. 2015; Haak et al. 2010, 2015; Hofmanová et al. 2016; Kivisild 2017; Lazaridis 2018; Lazaridis et al. 2014; Lipson et al. 2017; Mathieson et al. 2015, 2018; Mittnik et al. 2018; Skoglund et al. 2012). First farmers from different Greek locations have been assigned to hg G2a which has not been detected in populations inhabiting Europe before agriculturalists (Bollongino et al. 2013). Hg G2a has also been detected among ancient farmers from Central European sites in present-day Germany and Austria (Haak et al. 2010, 2015; Lipson et al. 2017; Mathieson et al. 2015, 2018). Hg G2a is present in agriculturalists from sites in modern Romania, Serbia, Macedonia, Croatia, Bulgaria, Czech Republic and Ukraine (Mathieson et al. 2015, 2018). In addition to hg G2a, hgs I2 (second to G2a) and R1 have been found among Neolithic individuals from sites in Germany, Poland, Czech Republic and Ukraine (Allentoft et al. 2015; Haak et al. 2015; Mathieson et al. 2015, 2018; Olalde et al. 2018). Furthermore, hgs such as C, E1, J2 and T1a have been detected in Neolithic farmers from the Balkans (Mathieson et al. 2018).

Agricultural practices in the Great Hungarian Plain started with the Starčevo–Körös–Criş (~8 kya) and Transdanubia Linearbandkeramik (~7 kya) cultures (Gamba et al. 2014; Haak et al. 2015; Szécsényi-Nagy et al. 2015). Ancient male individuals who represent several farming cultures from the

Carpathian Basin mostly belonged to chrY hgs G2a (Allentoft et al. 2015; Lipson et al. 2017; Szécsényi-Nagy et al. 2015) which is consistent with findings elsewhere in Europe. Additionally, hg I and its sub-lineage I2a have been found among ancient farmers of the region (Allentoft et al. 2015; Gamba et al. 2014; Lipson et al. 2017; Mathieson et al. 2015; Szécsényi-Nagy et al. 2015) as well as hgs H2-L281 and C1a2-V20 (C6-V20, names according to Karmin et al. 2015) (Gamba et al. 2014; Lipson et al. 2017; Mathieson et al. 2015).

The Late Neolithic and subsequent Bronze Age in Europe were marked by huge migrations of pastoralists from Pontic-Caspian Steppes (Yamnaya steppe herders) (Anthony 2007; Haak et al. 2015). These migrations may have also contributed to the dispersal of at least some Indo-European languages (Haak et al. 2015). Additionally, it has been shown that increase in Yamnaya ancestry may have generally coincided with the spread of pasture and cultivable land and decrease of broad-leaf forest in continental Europe (Racimo et al. 2020). During that time sub-clades of hgs R1a-M343 and R1b-M420 became common in Europe (Kivisild 2017). Although hg R1b-M343 has been detected both in HG and early farmers, this hg was not abundant (Kivisild 2017). Additionally, HG and early farmers did not belong to sub-clade R1b-M269 which also became common during the Bronze Age (Allentoft et al. 2015; Haak et al. 2015; Mathieson et al. 2015). Early Bronze Age male individuals dated to ~4.2–3.7 kya whose remains were discovered in burial sites in Central Europe (present-day Germany and Czech Republic) mainly belonged to sub-clades of hg R1b (Allentoft et al. 2015; Haak et al. 2015; Mathieson et al. 2015, 2018; Olalde et al. 2018). Similarly Early Bronze Age individuals from Poland carried R1b but also R1a sub-clades (Olalde et al. 2018). People who inhabited the Great Hungarian Plain during the Early Bronze Age integrated different components of archaeological cultures located in the Near East, the Steppe and Central Europe (Gamba et al. 2014; Olalde et al. 2018). Bronze Age individuals of the area have been shown to carry chrY hgs G2a, I2a, J2a and R1b (Allentoft et al. 2015; Gamba et al. 2014; Mathieson et al. 2018; Olalde et al. 2018). Among contemporary individuals from Estonia and Lithuania R1a lineages were the most abundant (Mittnik et al. 2018; Saag et al. 2017). The currently widespread hg N (N3a3'5-L392 according to Karmin et al. 2015) was first detected in the Kola Peninsula ~3.5 kya making it so far the earliest occurrence of said hg in Fennoscandia (Lamnidis et al. 2018).

The Bronze Age was followed by the Iron Age in the first millennium BC. One of the most significant Early Iron Age cultures was the Scythian culture which at its peak spread in the Eurasian Steppe from Central Asia to Ukraine and Central Europe (Järve et al. 2019; Juras et al. 2017). Scythian individuals from Ukraine mostly carried hg R1a (Järve et al. 2019). Interestingly, an Iron Age individual (radiocarbon age 830–980 BC) from Hungary carried hg N (Gamba et al. 2014). The first occurrence of haplogroup N in Estonia stems from the Iron Age, as three individuals dated to 40–770 BC were assigned to sub-clade N3-M46 (TAT) (Saag et al. 2019)

In the next historic period – the Middle Ages (around 500–1500 CE) – additional migrations took place and the population of Europe increased. Genetic analyses of Medieval individuals from German burial sites show that males mostly belonged to the currently widespread chrY hg R1b, however hg G2a was also detected (O'Sullivan et al. 2018; Rothe et al. 2015). The gene pool of populations residing in the Carpathian Basin was deeply influenced by the numerous migrations/invasions of different nomadic groups (described in chapter 2.5.1.1). So far, mostly aDNA from the Avars and Conquerors has been studied due to the prevalence of well excavated archaeological sites (Fóthi et al. 2020; Neparáczki et al. 2017, 2019). Avar males from sites dated to the 6–7th century have been mostly assigned to hg N3-M46 (name according to Karmin et al., 2015) (Neparáczki et al. 2019). Findings from the archaeological sites dated to the 9th–11th century reveal that predominant chrY hgs among the Conquerors were R1 and I (sub-lineage I2a) (Fóthi et al. 2020; Neparáczki et al. 2017, 2019). Another frequent hg was N3-M46 with Conqueror males belonging to several sub-lineages (Fóthi et al. 2020; Neparáczki et al. 2019). Although hg N3-M46 was present during 6–7th and 9–11th century populations, the composition of its sub-lineages was considerably different with N3a5-B197 found among the Avars and N3a2-M2118 and N3a4-Z1936 (hg names according to Karmin et al. 2015) among the Conquerors (Neparáczki et al. 2019). Interestingly, despite the prevalence of hg N during the Middle Ages, it is marginal among the modern Hungarian population (Bíró et al. 2015; Csányi et al. 2008; Pamjav et al. 2017; Völgyi et al. 2009).

3. AIMS OF THE STUDY

Current dissertation gives an in-depth characterization of the resolved phylogeny of human Y-chromosomal haplogroup N and its North Eurasia-specific geographic expanse. The phylogeny of the sub-clade N3a4 is further refined and the splits between sub-clades dated to explore the genetic link between Uralic-speaking Hungarians living in Central Europe and populations from the Ural Mountains and West Siberia – the putative homeland of contemporary Hungarians. Additional case study examines the patrilineal genetic composition of Mongolic speaking Kalmyks inhabiting the south-eastern region of the European Plain with the aim to reveal a potential genetic connection to other Oirat-speaking groups from Mongolia and Kyrgyzstan who have been spatially separated from Kalmyks for ~400 years.

In more general terms, the aim of the present study is to continue a long-lasting quest for reaching increasingly deeper insights into the genetic history of the Uralic-speaking peoples in the context of populations inhabiting northern Eurasia – a research direction that has started in Tartu more than 20 years ago. Continuous progress in it has been largely dictated by increasingly more sophisticated methods available for population geneticists. Indeed, all publications used in the current thesis make use of resequencing of whole Y-chromosomes and analysing large sample sets from many populations – an approach that became possible only recently.

REF I:

- To clarify the inner structure of hg N phylogeny and date the branching events using high coverage chrY sequencing data.
- To resolve the geographic patterns of hg N expansion with a focus on the most common in North Eurasia Y-chromosomal sub-clades – hgs N2a and N3.

REF II:

- To refine the phylogeny and estimate the divergence times within hg N3a4 by combining chrY high coverage sequencing data from Hungary with data from populations inhabiting Ural Mountains and West Siberia.
- To exemplify the geographical distribution of hg N3a4 and its sub-clades in a geographically extensive sample set.

REF III:

- To describe the patrilineal genetic composition of Kalmyk ethnic groups and their connection to Oirat-speaking populations in Mongolia.
- To characterise the genetic composition of Sart Kalmaks of Kyrgyzstan and the Tsaatans of Mongolia.
- To examine whether the members of the Khoshut dynasty in Kalmykia belong to the putative Genghis Khan chrY lineage.

4. MATERIALS AND METHODS

All publications included in this thesis are based on high-coverage full sequences of the Y chromosome. In the first study a total of 94 sequences including 43 novel sequences are used to determine successive branching events and date the phylogeny of Eurasian hg N concentrating on two most frequent sub-clades in northern Eurasia – N2a-P43 and N3-M46. Additionally, from a total set of 6521 samples belonging to 56 populations from Northern Eurasia, 1631 hg N Y chromosomes were detected and genotyped for 16 informative binary markers to explore in detail the phylogeography of hg N inner structure (Ref I).

The second study concentrates on Hungarians living in Central Europe and unlike their neighbours speak a Ugric language from the Uralic language family. The research refines the phylogeny of clade N3a4-Z1936 by using a total of 33 high-coverage Y-chromosomal sequences, including 5 Hungarian samples together with samples from the populations inhabiting regions close to the Ural Mountains and West Siberia. For phylogeographic analysis a total of 4985 samples from 46 populations were genotyped for the defining markers B535, B539 and its sub-lineages defined by B540 and B545 (Ref II).

The third study uses a total of 37 high-coverage chrY sequences of which 35 belong to hg C3-M217. To explore the paternal genetic connection between Kalmyks inhabiting the south-eastern region of the European Plain and other Oirat-speaking groups from Mongolia, Kyrgyzstan and China a total of 454 unrelated males from 11 Oirat populations were genotyped within the context of their neighbouring populations (Ref III).

A detailed description of the used DNA samples and conducted analyses are provided in the respective research articles and their supplementary materials.

The role of the author of the PhD thesis is described above, in chapter List of Original Publications.

All samples were obtained from unrelated volunteers after receiving informed consent in accordance with the guidelines of the ethical committees of the institutions involved.

5. RESULTS AND DISCUSSION

This section gives an overview of the three scientific publications that constitute the original part of the current thesis. The following is a compressed summary of the main results and conclusions. Details can be found in the references and their respective supplementary information.

5.1. The refined phylogeny and phylogeography of haplogroup N (Ref I)

5.1.1. Refined phylogeny of haplogroup N

One of the main objectives of this study was to refine and time-calibrate the phylogeny of hg N (Figure 1 Ref I). We used 54 sequences generated with Complete Genomics (Mountain View, California, USA) technology published in Karmin et al. (2015) and 43 novel samples sequenced at Gene By Gene using the "Big Y" service. To merge the novel sequences with the 54 published sequences, we obtained the sequence overlap between the two platforms and implemented the "re-mapping filter" as described in Karmin et al. (2015) After filtering the usable sequence length was cut down to 6.2 Mb. The total number of 94 chrY full sequences, including three hg O samples for phylogeny rooting, enabled us to determine, refine and date several lineages and sub-clades that were previously unresolved.

In Northern Eurasia, the most widespread hgs N3 and N2a diverged around 18 kya (95% Bayesian credible interval [BCI]=15.7–20.0 kya). We found that N2a sub-clades split around 9.3 kya (95% BCI = 7.8–10.9 kya). Hg N2a has two sub-clades: major N2a1-B523 and a minor sub-clade N2a2. Hg N3 sub-clades N3a'b and N3c split around 13 kya (95% BCI = 11.3–14.6 kya). Clade N3a'b has two sub-clades – a major N3a-L708 and less frequent N3b-B187. The deepest clade of N3a-L708 is N3a1-B211 that has a sister-clade N3a2'6-M2110 with a nested cladistic structure displaying a division of its major sub-clade N3a3'6-CTS6967 into 4 sub-clades in the mid-Holocene ~5 kya (95% BCI=4.4–5.7 kya). Additionally, we found a deep lineage N5 defined by marker B482 representing the first split within the phylogeny of hg N and showed deep sub-structure within subclade N4-F2930 (95% BCI=22.8–27.9 kya).

5.1.2. Phylogeography of N2a and N3

The broad Eurasian circum-Arctic spread of hg N has was described earlier (Rootsi et al. 2007). Using the new sequencing data results of the refined phylogeny of hg N and to reveal the pattern of the main sub-clades we compiled a

genotyping panel of over 6500 samples from 56 Eurasian populations. This panel included over 3500 new and 3000 samples from previous studies genotyped to a higher resolution (Altshuler et al. 2010; Balanovsky et al. 2008; Fedorova et al. 2013; Kushniarevich et al. 2013; Rootsi et al. 2007; Tambets et al. 2004).

Hgs N2a and N3 display a geographically structured distribution patterns within North Eurasia. Hg N2a has a very broad spread and majority of N2a males belong to N2a1-B523 clade (10–30%) that has a distribution area covering West and South Siberia, Taymyr Peninsula and Volga-Ural region, but this clade is absent from eastern parts of Siberia. The most abundant sub-lineage within hg N2a1-B523 splits to sub-clades defined by B525 and B478. Sub-clade N2a1-B478 is more frequent in western and southern parts of Siberia. Compared to its sister-clade, N2a1-B525 is scarce. Other N2a1 sub-clades defined by L1419 and B528 are respectively present mainly in northern and southern parts of VUR.

Hg N3-M46 sub-clades N3a1, N3a2, N3b and N3c are geographically distinct. Hg N3c is represented in our data with two individuals from Japan where hg N3 has been shown to have a frequency under 1% (Hammer et al. 2006; Rootsi et al. 2007), whereas it is absent from the the Eurasian mainland. Hg N3b-B187 is mostly found among South Siberian and Mongolian populations while its main sister-clade N3a-L708 is more common in other areas of North Eurasia. The oldest N3a sub-clade N3a1 (95% BCI=3.3–5.3 kya) is predominantly present in Volga-Uralic region and West Siberia. Hg N3a2'6-M2110 exhibits a noteworthy geographic and temporal spread. One of the main branches of N3a2'6 is N3a2-M2118 that overwhelmingly prevails in central Siberian populations residing in Yakutia (Yakuts, Evenks and Evens). Hg N3a2-M2118 can also be found at lower frequencies among Khanties and Mansis of West Siberia. Hg N3a2 and its sister-clade N3a3'6-CTS6967 split around 7 kya (95% BCI=6.1–8.3 kya).

Hg N3a3'6-CTS6967 reaches from East Siberia to Northeast Europe including the Baltics with subclades coalescing as recently as 5 kya (95% BCI=4.4–5.7) These subclades comprise of populations that are geographically and belong to several linguistic families meaning that the quite recent expansion of this linage is not connected to the much older split of these linguistic phyla. Though later diversifications within separated sub-clades took place in linguistically related populations (e.g. N3a5a-F4205 in Mongolic speakers). The most frequent N3a3'6 sub-clade among Baltic populations is N3a3-VL29 incorporating more than one third of Estonian, Latvian and Lithuanian Y chromosomes. This clade is also prevalent in Karelians and can be found among Sami and Finnish people albeit in smaller proportions. However, N3a3 is most abundant hg N sub-clade in Slavic-speaking Belarusians, Ukrainians and Russians with and exception of North European Russians, who more closely resemble other non-Slavic neighbouring populations of Northeast Europe such as Karelians, Vepsans and Finns, as most of their Y chromosomes belong to sub-clade N3a4-Z1936. This sub-clade is also a major N3 clade

among the Volga-Uralic populations especially Volga Tatars. Additionally, it is one of the main lineages in Bashkirs residing around southern parts of the Urals. In Northeast Europe N3a3-VL29 and N3a4-Z1936 distribution areas are partially overlapping. Other two N3a sublineages – N3a5-B197 and N3a6-B479 – are limited to East Eurasia. Hg N3a5a defined by marker F4205 can be found predominantly around Lake Baikal in the Mongolic-speaking populations of Buryats and Mongols. Its neighbouring clade N3a5b-B202 is found almost 5000 km away in East Siberia and Beringia particularly in Koryaks, Chukchis and Asian Eskimos. Hg N3a6 is specific to Nanais residing in the Russian Far East. The fact that all hg N3 Nanai males belong to N3a6 suggests a strong founder event. Similar founder events describe several Siberian populations such as Koryaks, Chuckis and Eskimos who boast great N3a5-B202 frequencies, and especially Yakuts due to high prevalence of N3a2.

5.1.3. Geographic spread of less common haplogroup N sub-clades

The deep N5-B482 lineage was unexpectedly found in a sample displaying mixed autosomal admixture profile (around 80% European/Mediterranean and 20% Southeast Asian). According to ISOGG Y-DNA tree (https://isogg.org) and YFull YTree (YTree v8.01.00, https://www.yfull.com/) this hg is named N2 (N2a, N2a1) and N, respectively, and defined by marker P189.2. Hg N5 (name according to nomenclature presented in Karmin et al. 2015 and Ilumäe et al. 2016) is reported by both ISOGG and YFull to be found mostly in Western Balkan countries such as Serbia, Croatia, Bosnia and Herzegovina and Slovakia and has some additional individuals from the UK, Montenegro, Romania, Russian and Turkey. However, because of sampling biases and scarcity of hg N5 this should be viewed with discretion. The rarity of N5 makes it difficult to detect, and more precise designation of sub-clades is difficult at this point.

Studies (Batini et al. 2015; Hu et al. 2015; Karmin et al. 2015) have shown that hg N-F2930 (N4-F2930 according to Karmin et al. 2015) can be found in South-east Asia, mainly in China. It is more abundant in the south where frequencies stay mostly around 10% or higher (Hu et al. 2015). Additionally, N-F2930 has also been reported in Cambodia and Vietnam (Hallast et al. 2014; Karmin et al. 2015; Poznik et al. 2013). This is also confirmed by ISOGG, where the designation for the hg is N1b-F2930, and YFull, where this hg is denoted as N-F2905. In this study we used 11 samples from China and Japan to detect first splits within this haplogroup. The first diversification in sub-clade N-F2930 happened ~16.2 kya (95% BCI=14.4–18.2 kya) – the deep inner structure retaining much older lineages is evident despite the small sample size.

5.2. Y-chromosomal haplogroup N lineages in Hungarians – a probable connection to the linguistically closest Uralic speakers near the Ural Mountain range (Ref II)

As explained in Chapter 2 (in particular 2.5.1.), there is a dilemma: Hungarian as a language belongs to the Ugric branch of the Uralic family of languages, whereas genetically the present-day Hungarians are the closest to their Central European neighbours and far away from the West Siberian Ob-Ugric Khanty and Mansi peoples. In Ref II we re-analyse aspects of the Y-chromosomal pools of Hungarians and the geographically distant populations inhabiting Ural Mountain region and West Siberia including Ugric speaking Khanty and Mansi

5.2.1. Phylogenetic tree of hg N3a4-Z1936 and the coalescence times of its sub-lineages

Using a total of 33 high-coverage Y chromosome full sequences, including 16 previously published (Ilumäe et al. 2016; Karmin et al. 2015; Wong et al. 2017) and 17 novel sequences, we reconstructed and refined the phylogeny of N3a4-Z1936. Five of the novel sequences were of Hungarian origin and two already published N3a2 samples were used as an outgroup (Figure 10).

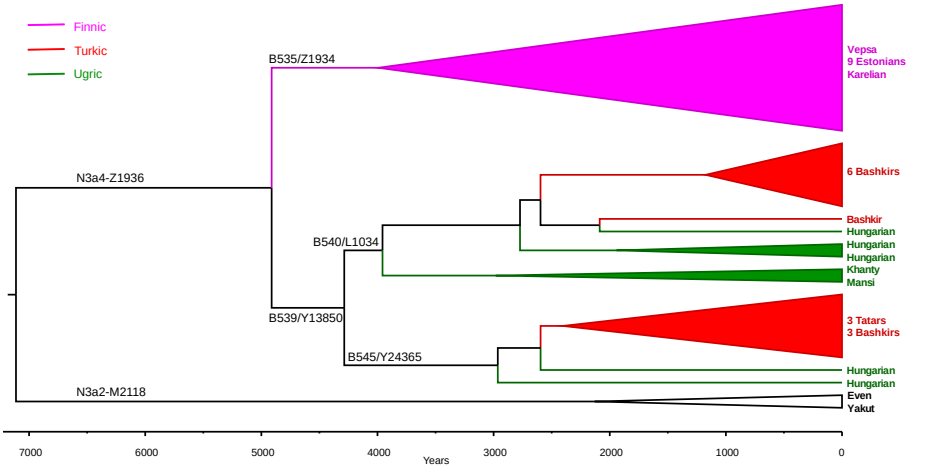

Figure 10. Reconstructed phylogeny of hg N3a4.
Two hg N3a2 samples were used as an outgroup. Sub-clades are presented as triangles (where possible) and coloured according to linguistic affiliations. Sub-clade and population names are displayed on the branches and tips, respectively.

The two major sub-clades of N3a4 defined by markers B535 and B539 diverged around 4.9 kya (95% BCI= 3.7–6.3 kya) and sub-clades of N3a4-B539 defined by markers B540 and B545 split 4.2 kya (95% BCI = 3.0–5.6 kya). Sub-clade N3a4-B540 has two sub-clades one of which consists of a Khanty and a Mansi sample and the other exclusively of Hungarian and Bashkir samples. The sister-clade of B540 (defined by B545) includes samples of Hungarians, Bashkirs and Tatars. Sub-clades defined by B540 and B545 diverged around 3.9 kya (95% BCI= 2.8–5.2 kya) and expanded almost simultaneously about 2.7–2.9 kya.

5.2.2. Geographic dispersal of N3a4 sub-lineages

To test the presence and frequencies of N3a4 sub-lineages in a broad sample set we used genotyping data of around 5000 Eurasian individuals including West Siberian Mansi and Khanty who are linguistically closest to Hungarians.

There is an apparent difference in the geographic distribution of hg N3a4 sub-clades as its frequency peaks in two separate areas – Northeast Europe where mostly N3a4-B535 sub-clade is found, and southern parts of the Ural Mountains where N3a4-B539 is common (Figure 3, Supplementary Table 3 Ref II). Hg N3a4-B535 has the highest frequency of 44% among Finns. However, it is also prevalent in Karelians, Saamis, Northern Russians (all around 20%) and Vepsians (32%). Northern Russians are an admixed population with a recorded similar genetic structure to their Finnic speaking neighbours (Balanovsky et al. 2008). The frequency of B535 diminishes to around 5% in Estonians and 1% in Latvians. In the east its frequency is 1–9% among Eastern European Russians and populations of the Volga-Ural region (Komis, Mordvins and Chuvashes). In Turkic speaking Bashkirs and Tatars residing in Southern Urals N3a4-B535 is mostly absent.

In comparison, the sister-clade of B535 defined by marker B539 is common among Bashkirs and can be found in Tatars (Figure 3, Supplementary Table 3 Ref II). N3a4-B539 is also present among Uralic speaking Mansis and Khanties of West Siberia and can be found in low frequencies (1–4%) among Hungarians of Central Europe. However, N3a4-B539 is almost completely missing from populations where N3a4-B535 lineages are prevalent such as Udmurts, Maris, Mordvins and Komis from the Volga-Uralic region and Finns, Vepsians, Saamis and Karelians from Northeast Europe. The two main sub-clades of B539 (defined by B540 and B545) have partially overlapping distributions. N3a4-B540 is most commonplace among different sub-groups of Bashkirs (up to 60%) but it has substantially lower frequencies among neighbouring Tatars (3–5%). B540 is also quite abundant in Khanties and Mansis (up to 27%) and can be found among Uralic speaking Nenets (7%) and Turkic speaking Dolgan populations (5%). Interestingly this lineage is present among Hungarians reaching the highest frequency (4%) in Sekler Hungarians who live in Transylvania, Romania. The presence of N3 (1%) has been shown

among Seklers previously (Csányi et al. 2008) but the low-resolution level makes it impossible to designate any sub-clades. B540 is absent from Uralic speaking Nganasans and Selkups. N3a4-B545, however, is much more localised being frequent (52%) in one Bashkir sub-group from Volga-Uralic region. While B545 is lacking in contemporary Mansis, it can be found in very small portions (<1%) among Hungarians, but it is absent in Sekler Hungarians.

5.2.3. N3a4-B539 – a link between Hungarians and linguistic relatives of Southern Urals and West Siberia

Present study refines the connection with high-resolution phylogenetic trees built from full sequences and genotypes a large sample set from geographically wide area. Both the hg N3a4 phylogenetic tree and genotyping data of hg N lineages show a connection between Hungarians and populations inhabiting Southern Urals and West Siberia. Hg N3a4 sub-clade defined by B539 is ubiquitous among Turkic-speaking Bashkirs and has a substantial frequency among Ob-Ugric Khanties and Mansis of West Siberia. Genotyping confirms that this lineage is present among Hungarians, whose hg N3a4 Y-chromosomes – similarly to Bashkir and Tatar Y-chromosomes – are present in both hg N3a4-B539 sub-clades (defined by markers B540 and B545). The core area of hg N3a4-B539 is located in proximity to the Urals, but it is also present at low frequency in Central European Hungarians with an absence among neighbouring Indo-Europeans (Figure 3 Ref II) – this noteworthy pattern is difficult to explain by gradual frequency cline and without an assumption of a migration of people.

To test whether the presence of hg N among Sekler Hungarians could be caused by random drift, selection based resampling approach was implemented. Three models (details in Material and Methods section of Ref II) were implemented which considered the observed frequencies of hg N3a4-B539 in Hungarian Seklers (4%), neighbouring Indo-European populations (0%) and Southern Ural/West Siberian populations (13%) and evaluated the possibility of a genetic contribution from the Ural/West Siberian populations to Hungarian Seklers. Statistically most supported model (model C) demonstrated that drift alone cannot explain this frequency pattern and that Sekler Hungarians display a genetic contribution from Siberia that is missing in neighbouring Indo-Europeans.

Although the frequency of hg N3a4-B539 is marginal among present-day Hungarians, it is possible that it was higher in ancient Magyars. The Z1936 lineage was found in 5 individuals out of 19 (26,3%) in the archaeologically richest Hungarian late 9th-early 10th century cemeteries from Upper-Tisza and Middle-Tisza basin regions (Fóthi et al. 2019, 2020). Notably, this frequency is quite similar to marker Z1936 derived state frequencies among various contemporary Khanty, Mansi and Bashkir groups. Additionally, it has been shown that two ancient Hungarian N3a4-Z1936 samples from these

cemeteries deduced to clade B545 according to genotyping results (Fóthi et al. 2019). We added these samples to the STR network analysis (Supplementary Figure 2 Ref II), and the results showed a closer relationship to contemporary Bashkir and Tatar samples than to present-day Hungarians.

5.3. Paternal genetic relatedness of Kalmyks in the southeast European Plain and Oirats of Western Mongolia (Ref III)

Kalmyks are the only Mongolic-speaking people settled down in Europe. Chapter 2.5.2. gives an overview about their place among different Mongolic speakers, their clan structure and published genetic data. Below new results are presented and discussed.

5.3.1. Reconstructed phylogeny of hg C3-M217 and the distribution of its sub-clades among Mongol-speaking Kalmyk and Oirat-Mongol populations

In this study we genotyped a total of 454 samples from 33 haplogroups. Different hg C3 sub-clades are most abundant (>50%) paternal lineages among studied populations (Kalmyk Derbet, Kalmyk Torgut, Kalmyk Buzava, Kalmyk Khoshut, Sart Kalmak, Xinjiang Kalmyk, Mongol Torgut, Mongol Khoshut, Mongol Derbet, Mongol Tsaatan, Tozhu Tuvan).

We reconstructed the phylogeny of hg C3-M217 based on 10 novel and 27 published (Karmin et al. 2015) Y-chromosome whole sequences (Figure 11) focusing on two sub-branches C3f1 and C3c'i which are the most common among the Kalmyk population. The phylogeny of hg C3c'i sub-clade C3c1-M77 displays two sub-branches: sub-branch C3c1a-Z40439 has been formerly described and is common among Tungusic-speaking Evens and Evenks and Mongolic-speaking Buryats and Mongols (Karmin et al. 2015). The sub-branch C3c1b-F6379 is novel with a TMRCA of ~1.5 thousand years (95% BCI = 1.1–2.1 kya). We show that sub-lineages of hg C3 form the major shared component between the various sub-populations of Kalmyks and Mongolian Oirat sub-groups. Genotyping data reveals that the newly described C3c1b is the most common sub-clade among Mongolic Kalmyks and Oirats covering ~40% of male lineages. Although this hg is also present at low incidence in Turkic-speaking Tsaatans and Oirat-speaking Sart Kalmaks (frequencies are 4.3% and 4.9%, respectively). When Tsaatans, Tozhu Tuvans, and Sart Kalmaks are excluded, the share of hg C3 reaches 62.5%. Similarly, the supposed Khoshut ruler from Kalmykia with a possible link to Genghis Khan and his brother belongs to a lineage in sub-clade C3c1-M77. However, since C3c is extremely widespread in Central Asia, Mongolia and other parts of East

Asia, the connection to Genghis Khan dynasty cannot be verified and the finding does not provide additional information concerning the Khan dynasty chrY lineage.

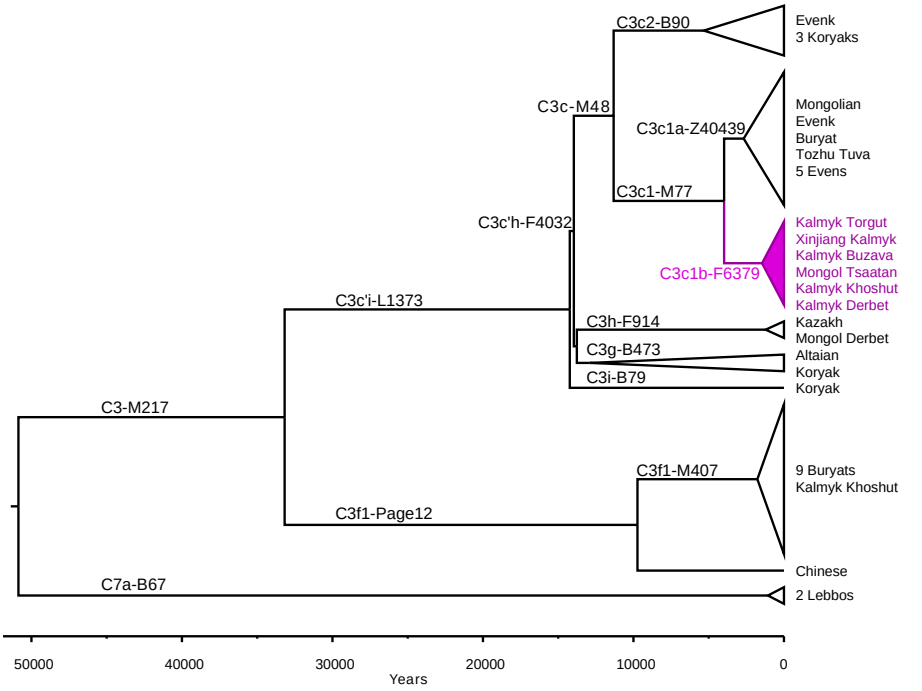

Figure 11. The phylogenetic tree of hg C3.
Newly defined C3 sub-clade C3c1b-F6379 is indicated in purple. Two C7a samples were used as an outgroup. Sub-clades are presented as triangles (where possible). Sub-clade names on the branches and population identifiers on the tips.

5.3.2. Presence of haplogroup N3 sub-clades and dispersal of other paternal haplogroups in Kalmyks and other Oirat groups

Although hg C3 sub-clades are most common among different Oirat populations, the incidence of other haplogroups varies (Figure 1 Ref III). Other Y-chromosomal haplogroups present in Kalmyks and Oirat-speakers are mostly of Siberian or East Asian origin such as N, O, Q, R1a2-Z93 derivates, R2.

For example, hg N is more frequent in Kalmyk Khoshuts (32.1%) compared to the other populations included in this study. Hg N3a2-M2118, common in Yakutia (Central Siberia) (Ilumäe et al., 2016), is carried by 21.4% of Kalmyk Khoshuts, who also have 7.1% of hg N3a5a-F4205, commonly found in Buryats and Mongols (Ilumäe et al. 2016). Among the Tsaatans of Mongolia, however, hg N3a5a-F4205 is a major Y-chromosome lineage constituting more than 50% of the paternal genepool. Additionally, this lineage also has a

high frequency in Tozhu Tuvans (15.2%), Mongol Derbets (12.5%), Mongol Torguts (10.6%) and Mongol Khoshuts (5.6%). Hg N lineages like N2a1, N3b and N3a3'6-CST696 are also present in Kalmyks and other Oirat-speakers albeit in lower frequencies. For example, a sub-clade of hg N2a1 defined by marker B478 is present in 13% of Tozhu Tuvan Y-chromosomes, whereas it is less common or absent among other Oirat-speaking groups (0–5%). Furthermore, Tozhu Tuvans are the only studied population that has a marginal proportion of N3b (around 2%).

Another frequent haplogroup among Oirat-speaking groups is hg O. For example, hg O2a2-P201 that is prevalent in China and Southeast Asia has a considerable frequency among Kalmyk Khoshuts (~14%) and Mongol Derbets (~7%). Kalmyk Torguts have a substantial amount of hg O2'5 (~17%) but we were not able to further resolve this haplogroup. Hgs O2'5 and O2a2 are also frequent in a limited Kalmyk sample from the Chinese autonomous region of Xinjiang but due to the small sample size (12) these results should be considered with caution.

Kalmyks and other Oirats also carry sub-lineages of hg R such as R1a2-Z93 and R2. It should be noted that Sart Kalmaks deviate from other studied populations due to high frequency of hg R and its sub-group R1a-Z2125 (~30%) that is common in Kyrgyzstan and Afghan Pashtuns (Underhill et al., 2015). This may point to a male mediated gene flow between the neighbouring Sart-Kalmaks and Kyrgyz. In addition, more than 11% of Sart-Kalmak Y-chromosomes belong to R1b13-M478 that also has a noticeable frequency among Kalmyk-Derbets (~7%). The latter, along with Kalmyk Buzavas, also have a significant amount of R2a-M124 (>18% and >11%, respectively).

The Tsaatans of Mongolia who, in addition to high frequency of hg N3a5a (52%), also have a substantial amount of hg Q2a-M25 (43.5%). The fact that these two hgs constitute around 96% of the paternal gene pool of Mongolian Tsaatans indicates a founder effect followed by genetic isolation, typical of a small population of taiga reindeer herders. The same lineages dominate in the Tozhu Tuvans with hg Q2a-M25 contributing the most by having a frequency of 50.0%. This provides genetic proof of inter-marriage between these two groups. Previously published Y-chromosomal genotyping data shows that Mongolian Tsaatans and Tozhu Tuvans are also similar to Tuvans (Kharkov et al. 2013) pointing to a degree of common paternal origin between the three populations.

Despite the almost 400 years of geographical separation, the distribution of Y-chromosomal haplogroups among Kalmyk sub-groups and Oirats of Western Mongolia is quite similar. According to pairwise genetic distances (Fst) based on chrY hg frequencies, Mongol Tsaatan, Tozhu Tuvan and Sart Kalmak display significant divergence from the remaining populations studied. Repeating the PCA based on chrY haplogroup frequencies without these three groups highlights the displacement of Kalmyk Derbets, which is driven by the high frequency of haplogroup R2a among them. Other strong drivers of differentiation are hg N3a2-M2118 present in Kalmyk Khoshuts, and hg

O2’5-M268 with the highest level among Kalmyk Torguts. The AMOVA results (Supplementary Table S4 Ref III) validate the arrangement seen on the PC plot (Figure 1b Ref III) (FCT=0.15670, P=0.00293) showing a low (FCT=0.01771) and insignificant (P =0.13196) level of differentiation among the geographically distant groups. This may be because the paternal genetic structure of Oirat-speaking Mongol sub-populations had developed in the territory of Western Mongolia and has not undergone significant changes after the split, neither among the Kalmyk people nor in other Oirats of Inner Asia.

CONCLUSIONS

- Haplogroup N is a widespread chrY lineage in Northern Eurasia. The refined hg N phylogeny has disclosed several previously undescribed sub-clades. Subsequent genotyping of defining markers outlines the previously homogenous geographic structure into distinctive hg N lineages.
- One of the most widespread hg N sub-lineages is N2a-P43.
- The inner structure of the largest hg N2a sub-group N2a1 contains three sub-clades instead of previously known two.
- Hg N2a males overwhelmingly belong to hg N2a1-B523 which is present in West and South Siberia, Taymyr Peninsula and the Volga-Uralic region.
- The other prevalent hg in Northern Eurasia is N3-M46.
- Haplogroup N3 has two sub-clades – N3a'b and N3c. Sub-clade N3a splits into N3a1-B211 and N3a2'6-M2110, the latter of which divided almost simultaneously into 2 sub-lineages.
- Hg N3a3'6 has a wide geographic spread and almost simultaneous expansion of sub-lineages starting ~5kya into multiple distinct geographical areas of North Eurasia.
- The distribution pattern of hg N3a3'6 sub-lineage N3a4 has two apparent frequency peaks – Northeast Europe (sub-clade N3a4-B535) and southern parts of the Ural Mountains (sub-clade N3a4-B539).
- Sub-clade N3a4-B539 forms a genetic link between Hungarians residing in Central Europe and geographically distant populations of West Siberia and southern Urals including linguistic relatives.
- Sub-clade N3a4-B539 frequency among present-day Hungarians is marginal being the highest ~4% among Sekler Hungarians and staying at ~1% in other Hungarian sub-groups.
- Hg C3-M217 sub-lineages make up over 40% of Kalmyk male lineages.
- The phylogeny of hg C3-M217 displays two sub-branches C3f1 and C3c'i. One of the hg C3c'i sub-lineages is the novel C3c1b-F6379 that is specific to the Kalmyks.
- Although hg C3 is the most common hg among Mongolic Kalmyks and other Oirat-speakers, additional hgs mostly of Siberian or East Asian origin (hg N, O, Q, R1a2-Z93 derivates, R2) are present in varying frequencies.

SUMMARY IN ESTONIAN

Ülevaade Põhja-Euraasias laialt levinud Y-kromosoomaalse haplogrupi N fülogeneesist ja fülogeograafiast ning Euroopa kahe keeleliselt erandliku rahva – ungarlaste ja kalmõkkide – populatsiooniuuringud

Inimkonna evolutsioon ja rahvaste demograafiline ajalugu ehk protsessid, mis on mõjutanud rahvastiku muutumist läbi aja, on alati inimesi huvitanud. Need sündmused on mõjutanud ka inimese põlvest-põlve edasi kanduvat genoomi, mis seetõttu peidab endas inimese evolutsioonilist ning demograafilist ajalugu. Populatsioonigeneetiline teadussuund on kasutanud geneetilisi andmestikke rahvastikusündmuste uurimisel juba aastakümneid, kuid just praegu viivad kiired tehnoloogilised arengud teadusharu hüppeliselt edasi.

Üks kõige olulisemaid edasiminekuid viimase 15 aasta jooksul on uue lühikeste lugemete järjestusanalüüsi (LLJ) väljatöötamine. Antud tehnoloogia tulemusel kasvas väga oluliselt võime analüüsida geneetilisi andmeid. Tänu LLJ meetodile suudame palju paremini kogu tuumagenoomi järjestada ning samuti on võimalik analüüsida ammusurnud indiviidide genoomset DNAd ehk vana DNAd. Niisamuti on nüüd palju enam võimalusi vaid ema- või isaliini pidi päranduvate ehk uniparentaalsete genoomi osade, mitokondri DNA (mtDNA) ning Y-kromosoomi DNA (Y-DNA) analüüsiks. Kuigi kogu genoomi ja vana DNA analüüs annavad põhjalikumat ja mitmekülgsemat informatsiooni inimkonna evolutsiooni ning demograafilise ajaloo kohta, võimaldavad uniparentaalsete genoomi osade omadused analüüsida demograafilise ajaloo mees- ja naisliinidele omaseid aspekte, mis oluliselt rikastab meie võimet mõista inimkonna demograafilist ajalugu selle kogu rikkuses.

Käesolev töö keskendub inimese Y-kromosoomi analüüsile. Y-kromosoom on haploidne kromosoom, esineb ainult meestel ja suurem osa kromosoomist ei läbi meiootilist ristsiiret. Ainult kromosoomi otsad vahetavad rekombinatsiooni käigus X-kromosoomiga geneetilist materjali. Seetõttu on muutused, mis Y-kromosoomis aset leiavad, tingitud põhiliselt aja jooksul kuhjunud mutatsioonidest. Üheks informatiivsemaks uurimisobjektiks on ühe-aluspaarilised punktmutatsioonid ehk SNV-d (ingl k *single nucleotide variant*). Y-kromosoom pärandub ainult meesliini pidi (isalt pojale), kandes endas ellujäänud liinide infot ühisest eellasest kuni tänapäeva indiviidideni. SNV-de alleelivariantide kombinatsioonide põhjal saame määrata Y-kromosoomid fülogeneetilistesse gruppidesse ehk haplogruppidesse (hg), mis omakorda koonduvad kõiki isaliine siduvaks fülogeneetiliseks puuks. Kõigi nende tunnuste tõttu on Y-kromosoomi kasutatud meesliinide uurimiseks rohkem kui 20 aastat. Algusaegadel põhinesid teadusuuringud vähesel hulgal punktmutatsioonidel, kuid meetodite arenedes lisandus aina uusi informatiivseid markereid. Uus LLJ tehnoloogia võimaldab vaid teadaolevate punktmutatsioonide asemel analüüsida Y-kromosoomi DNA järjestust hüppeliselt

suuremal hulgal kui varem. Sellega kaasneb aina rohkemate markerite leidmine ning omakorda isaliinide põlvnemist rekonstrueerivate fülogeneesipuude täpsustamine. Klassikaline fülogeograafiline analüüs – demograafiliste sündmuste rekonstrueerimine lähtuvalt isaliinide geograafilisest levikust ja fülogeneetilistest seostest, on seetõttu saanud oluliselt nüansseeritumaks. Antud käsitlusviis on laialdaselt kasutusel nii Y-kromosoomi globaalse kui ka populatsioonipõhise varieeruvuse analüüsil.

Käesolevas töös on üheks eesmärgiks täiendada Põhja-Euraasias laialt levinud haplogrupi N fülogeneesi sisestruktuuri, kasutades selleks 94 Y-kromosoomi täisjärjestust. Täiendavalt kirjeldatakse hg N-i ja selle põhiliste alamklaadide geograafilist levikut Euraasia populatsioonides.

Teiseks sihiks on uurida geneetilist seost Kesk-Euroopas elavate uurali keeli kõnelevate ungarlaste ja geograafiliselt kaugete Lääne-Siberis ning Uurali mäestikus elavate uurali keeli kõnelevate populatsioonide, sealhulgas lähimate keelesugulaste vahel. Lääne-Siberit ning Uurali mäestiku ümbrust on pikalt peetud uurali keelte kõnelejate, kaasa arvatud ungarlaste, algkoduks. Teada on, et tänapäeva ungarlased on geneetiliselt väga sarnased oma geograafiliste lähinaabritega, kes kõnelevad indoeuroopa keeli. Ungarlaste side oma uurali keeli kõnelevate keelesugulastega, eriti just ugri keeli rääkivate rahvastega, on nii tuumagenoomi, mtDNA kui Y-DNA kontekstis pea olematu. Siiski leidub väike hulk Y-DNA liine N haplogrupist, mis viitavad seosele Lääne-Siberis elavate populatsioonidega. Selle seose selgitamiseks rekonstrueeriti 33 Y kromosoomi täisejärjestuse põhjal N3a4-Z1963 fülogeneesipuu, kasutades sealjuures esmakordselt ungarlaste Y-kromosoomi täisjärjestusi. Lisaks kirjeldab antud uuring N3a4 alamklaadide geograafilist levikut uuritud populatsioonides.

Euroopas ainsana mongoli keelt kõnelevad kalmõkid on budistlik rahvas, kelle Oiratitest esivanemad siirdusid umbes 400 aastat tagasi oma algkodust Mongooliast Ida-Euroopa lauskmaa kaguossa. Töö eesmärgiks on võrrelda, naaberrahvaste taustal, kalmõkkide isaliine nende Mongoolias, Kõrgõzstanis ja Hiinas elavate keelesugulastega, lähtudes sealjuures ka tänini säilunud teadmisest Oiratite klannistruktuurist. Kirjeldamaks ajalise ning geograafilise lahusoleku mõju kalmõkkide geneetilisele varieeruvusele, rekonstrueeritakse laialt levinud Y-kromosoomi haplogrupi C3-M217 fülogeneetiline puu. Täiendavalt selgitatakse nii hg C3 kui ka teiste haplogruppide esinemissagedusi kalmõkkidel ja teistel oiratidel.

Doktoritöö põhiliste tulemuste kokkuvõte:

- Y-kromosomaalne haplogrupp N on laialt levinud erinevates Põhja-Euraasia populatsioonides ning tekkis Kagu-Aasia mandriosas, kus vanaDNA andmete kohaselt oli hg N levinud umbes 6500 aastat tagasi.
- Põhjalik ja kõrge lahutusastmega hg N fülogeneetiline puu paljastab mitu seni kirjeldamata alamklaadi.
- Kaks kõige levinumat N alamliini, N2a ning N3, lahknesid umbes 18000 aastat tagasi.

- N2a kõige sagedasema alamklaadi N2a1-B523, mis on levinud Lääne- ja Lõuna-Siberis, Taimõri poolsaarel ning Volga-Uurali piirkonna rahvastel, sisestruktuur jaguneb varasemalt arvatud kahe klaadi asemel kolme klaadi.
- Hg N3 koosneb kahest alamliinist: N3a'b ja N3c.
- Hg N levikumustri kõige silmapaistvam aspekt on N3a3'6 klaadi alamliinide lai geograafiline ulatus ja nelja alamklaadi kiire jagunemine umbes 5000 aastat tagasi.
- Hg-l N3a4 on kaks sagedusharipunkti. Üks neist asub Kirde-Euroopas ja teine Uurali mäestiku lõunaosas.
- N3a4 on üks väheseid geneetilisi ühendusi ungarlaste ning teiste uurali keeli kõnelevate rahvaste vahel. Kuigi N3a4 sagedus tänapäeva ungarlastel on marginaalne, oli see ilmselt vanadel madjaritel kõrgem, millele viitavad ka vana DNA uuringud.
- Kui välja jätta Kõrgõzstanis elavad sart-kalmõkid, siis hg C3 moodustab kalmõkkide ning teiste oiratide isaliinide geenitiigist enam kui 60%.
- Haplogrupi C3 fülogeneesipuu paljastab sisestruktuuris uue alamliini C3c1b-F6379, mis on oiratidel sh kalmõkkidel levinuim alamklaad (~40%).

REFERENCES

Abilev S, Malyarchuk B, Derenko M, Wozniak M, Grzybowski T, and Zakharov I. 2012. The Y-Chromosome C3* Star-Cluster Attributed to Genghis Khan's Descendants Is Present at High Frequency in the Kerey Clan from Kazakhstan. *Human Biology* 84(1):79–89.

Akbulatov IM. 1988. Керамика Таптыковского Городища Эпохи Раннего Средневековья. Pp. 31–35 in *Проблема Древних Угров На Южном Урале*. Ufa: БНЦ УрО АН СССР

Al-Zahery N, Pala M, Battaglia V, Grugni V, Hamod MA, Kashani B, et al. 2011. In Search of the Genetic Footprints of Sumerians: A Survey of Y-Chromosome and MtDNA Variation in the Marsh Arabs of Iraq. *BMC Evolutionary Biology* 11(1).

Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A, et al. 2003. Y-Chromosome and MtDNA Polymorphisms in Iraq, a Crossroad of the Early Human Dispersal and of Post-Neolithic Migrations. *Molecular Phylogenetics and Evolution* 28(3):458–72.

Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J, et al. 2015. Population Genomics of Bronze Age Eurasia. *Nature* 522(7555):167–72.

Altshuler DL, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, et al. 2010. A Map of Human Genome Variation from Population-Scale Sequencing. *Nature* 467(7319):1061–73.

Alvarez L, Ciria E, Marques SL, Santos C, and Aluja MP ila. 2014. Y-Chromosome Analysis in a Northwest Iberian Population: Unraveling the Impact of Northern African Lineages. *American Journal of Human Biology : The Official Journal of the Human Biology Council* 26(6):740–46.

Anonymus. 2010. *The Deeds of the Hungarians. Master Roger's Epistle to the Sorrowful Lament upon the Destruction of the Kingdom of Hungary by the Tartars*. edited by J. M. Bak and M. Rady. Budapest; New-York: Central European University Press

Anthony DW. 2007. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton: Princeton University Press

Ashilova DO. 1976. *Ethnic Anthropology of the Kalmyks*. Elista: Kalmyk Book Publishing House

Avise JC, Arnold J, Ball, R.M., Bermingham E, Lamb T, Neigel JE, et al. 1987. Intraspecific Phylogeography: The Mitochondrial DNA Bridge between Population Genetics and Systematics. *Annual Review of Ecology and Systematics* 18:489–522.

Bachtrog D. 2008. The Temporal Dynamics of Processes Underlying Y Chromosome Degeneration. *Genetics* 179(3):1513–25.

Bachtrog D. 2013. Y-Chromosome Evolution: Emerging Insights into Processes of Y-Chromosome Degeneration. *Nature Reviews Genetics* 14(2):113–24.

Balanovsky O. 2017. Toward a Consensus on SNP and STR Mutation Rates on the Human Y-Chromosome. *Human Genetics* 136(5):575–90.

Balanovsky O, Dibirova K, Dybo A, Mudrak O, Frolova S, Pocheshkhova E, et al. 2011. Parallel Evolution of Genes and Languages in the Caucasus Region. *Molecular Biology and Evolution* 28(10):2905–20.

Balanovsky O, Gurianov V, Zaporozhchenko V, Balaganskaya O, Urasin V, Zhabagin M, et al. 2017. Phylogeography of Human Y-Chromosome Haplogroup Q3-

L275 from an Academic/Citizen Science Collaboration. *BMC Evolutionary Biology* 17(Suppl 1):1–15.

Balanovsky O, Rootsi S, Pshenichnov A, Kivisild T, Churnosov M, Evseeva I, et al. 2008. Two Sources of the Russian Patrilineal Heritage in Their Eurasian Context. *American Journal of Human Genetics* 82(1):236–50.

Balanovsky O, Zhabagin M, Agdzhoyan A, Chukhryaeva M, Zaporozhchenko V, Utevska O, et al. 2015. Deep Phylogenetic Analysis of Haplogroup G1 Provides Estimates of SNP and STR Mutation Rates on the Human Y-Chromosome and Reveals Migrations of Iranic Speakers. *PLoS ONE* 10(4).

Balaresque P, Bowden GR, Adams SM, Leung HY, King TE, Rosser ZH, et al. 2010. A Predominantly Neolithic Origin for European Paternal Lineages. *PLoS Biology* 8(1).

Balaresque P, Poulet N, Cussat-Blanc S, Gerard P, Quintana-Murci L, Heyer E, et al. 2015. Y-Chromosome Descent Clusters and Male Differential Reproductive Success: Young Lineage Expansions Dominate Asian Pastoral Nomadic Populations. *European Journal of Human Genetics* 23(10):1413–22.

Balinova NV. 2015. Anthropometric Study of Sub-Ethnic Groups of the Kalmyks [in Russian]. *Bulletin of the Kalmyk Institute for Humanities RAS* 21:93–101.

Ballantyne KN, Goedbloed M, Fang R, Schaap O, Lao O, Wollstein A, et al. 2010. Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Rorensic Implications. *American Journal of Human Genetics* 87(3): 341–53.

Barać L, Peričić M, Klarić IM, Rootsi S, Janićijević B, Kivisild T, et al. 2003. Y Chromosomal Heritage of Croatian Population and Its Island Isolates. *European Journal of Human Genetics* 11(7):535–42.

Barbieri C, Hübner A, Macholdt E, Ni S, Lippold S, Schröder R, et al. 2016. Refining the Y Chromosome Phylogeny with Southern African Sequences. *Human Genetics* 135(5):541–53.

de Barros Damgaard P, Martiniano R, Kamm J, Víctor Moreno-Mayar J, Kroonen G, Peyrot M, et al. 2018. The First Horse Herders and the Impact of Early Bronze Age Steppe Expansions into Asia. *Science* 360(6396).

Batbayar K and Sabitov ZM. 2012. The Genetic Origin of the Turko-Mongols and Review of the Genetic Legacy of the Mongols. Part 1: The Y-Chromosomal Lineages of Chinggis Khan. *The Russian Journal of Genetic Genealogy* 4(2):1–8.

Batini C, Ferri G, Destro-Bisol G, Brisighelli F, Luiselli D, Sánchez-Diz P, et al. 2011. Signatures of the Preagricultural Peopling Processes in Sub-Saharan Africa as Revealed by the Phylogeography of Early y Chromosome Lineages. *Molecular Biology and Evolution* 28(9):2603–13.

Batini C, Hallast P, Zadik D, Delser PM, Benazzo A, Ghirotto S, et al. 2015. Large-Scale Recent Expansion of European Patrilineages Shown by Population Resequencing. *Nature Communications* 6(May 2014).

Battaglia V, Fornarino S, Al-Zahery N, Olivieri A, Pala M, Myres NM, et al. 2009. Y-Chromosomal Evidence of the Cultural Diffusion of Agriculture in Southeast Europe. *European Journal of Human Genetics* 17(6):820–30.

Battaglia V, Grugni V, Perego UA, Angerhofer N, Gomez-Palmieri JE, Woodward SR, et al. 2013. The First Peopling of South America: New Evidence from Y-Chromosome Haplogroup Q. *PLoS ONE* 8(8).

Behar DM, Saag L, Karmin M, Gover MG, Wexler JD, Sanchez LF, et al. 2017. The Genetic Variation in the R1a Clade among the Ashkenazi Levites' y Chromosome. *Scientific Reports* 7(1):1–11.

Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho TJ, et al. 2014. Mammalian y Chromosomes Retain Widely Expressed Dosage-Sensitive Regulators. *Nature* 508(7497):494–99.

Benazzi S, Fornai C, Bayle P, Coquerelle M, Kullmer O, Mallegni F, et al. 2011. Comparison of Dental Measurement Systems for Taxonomic Assignment of Neanderthal and Modern Human Lower Second Deciduous Molars. *Journal of Human Evolution* 61(3):320–26.

Bergström A, Nagle N, Chen Y, McCarthy S, Pollard MO, Ayub Q, et al. 2016. Deep Roots for Aboriginal Australian y Chromosomes. *Current Biology* 26(6):809–13.

Bhowmick BK, Satta Y, and Takahata N. 2007. The Origin and Evolution of Male-Specific Gene Families in the Human Y Ampliconic Region. *Biosystems* 441–50.

Bíró A, Fehér T, Bárány G, and Pamjav H. 2015. Testing Central and Inner Asian Admixture among Contemporary Hungarians. *Forensic Science International: Genetics* 15:121–26.

Birtalan A. 2012. *Oirad and Kalmyk Linguistic Essays*. edited by K. S. Ernő and S. Gábor. Budapest, Hungary: ELTE Eötvös Kiadó

Boattini A, Sarno S, Bini C, Pesci V, Barbieri C, De Fanti S, et al. 2016. Mutation Rates and Discriminating Power for 13 Rapidly-Mutating Y-STRs between Related and Unrelated Individuals. *PLoS ONE* 11(11):1–14.

Boattini A, Sarno S, Mazzarisi AM, Viroli C, De Fanti S, Bini C, et al. 2019. Estimating Y-Str Mutation Rates and Tmrca Through Deep-Rooting Italian Pedigrees. *Scientific Reports* 9(1):1–12.

Boldog Z, Petkes Z, Sudár B, and Türk A. 2015. From the Ancient Homelands to the Carpathians – From the Finds to the Apparel. *Hungarian Archaeology E-Journal* Autumn.

Bollongino R, Nehlich O, Richards MP, Orschiedt J, Thomas MG, Sell C, et al. 2013. Stone Age Central Europe. *Science* 342(October):479–81.

Bougdaeva S and Isaacs R. 2018. Nomads under Arrest: The Nation-Building and Nation-Destroying of Kalmyk Nomads in Russia. *Communist and Post-Communist Studies* 51(4):375–85.

Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, et al. 2009. Genetic Discontinuity between Local Hunter-Gatherers and Central Europe's First Farmers. *Science* 326(5949):137–40.

Brandt G, Szécsényi-Nagy A, Roth C, Alt KW, and Haak W. 2015. Human Paleogenetics of Europe – The Known Knowns and the Known Unknowns. *Journal of Human Evolution* 79:73–92.

Bugoye FC, Mulima E, and Misinzo G. 2018. Analysis of Mutation Rate of 17 Y-Chromosome Short Tandem Repeats Loci Using Tanzanian Father-Son Paired Samples. *Genetics Research International* 2018.

Burgarella C and Navascués M. 2011. Mutation Rate Estimates for 110 Y-Chromosome STRs Combining Population and Father-Son Pair Data. *European Journal of Human Genetics* 19(1):70–75.

Busby GBJ, Brisighelli F, Sánchez-Diz P, Ramos-Luis E, Martinez-Cadenas C, Thomas MG, et al. 2012. The Peopling of Europe and the Cautionary Tale of Y Chromosome Lineage R-M269. *Proceedings of the Royal Society B: Biological Sciences* 279(1730):884–92.

Cai X, Qin Z, Wen B, Xu S, Wang Y, Lu Y, et al. 2011. Human Migration through Bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum Revealed by Y Chromosomes. *PLoS ONE* 6(8).

Capelli C, Brisighelli F, Scarnicci F, Arredi B, Caglia' A, Vetrugno G, et al. 2007. Y Chromosome Genetic Variation in the Italian Peninsula Is Clinal and Supports an Admixture Model for the Mesolithic-Neolithic Encounter. *Molecular Phylogenetics and Evolution* 44(1):228–39.

Charlesworth B. 1978. Model for Evolution of Y Chromosomes and Dosage Compensation. *Proceedings of the National Academy of Sciences of the United States of America* 75(11):5618–22.

Charlesworth B. 1991. Evolution of Sex Chromosomes. *Nature* 147(3739):809.

Charlesworth B and Charlesworth D. 2000. The Degeneration of Y Chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 355(1403): 1563–72.

Chiaroni J, Underhill PA, and Cavalli-Sforza LL. 2010. Erratum: Y Chromosome Diversity, Human Expansion, Drift, and Cultural Evolution (Proceedings of the National Academy of Sciences of the United States of America (2009) 106, 48 (20174-20179) DOI: 10.1073/Pnas.0910803106). *Proceedings of the National Academy of Sciences of the United States of America* 107(30):13556.

Cinnioğlu C, King R, Kivisild T, Kalfoğlu E, Atasoy S, Cavalleri GL, et al. 2004. Excavating Y-Chromosome Haplotype Strata in Anatolia. *Human Genetics* 114(2):127–48.

Claerhout S, Vandenbosch M, Nivelle K, Gruyters L, Peeters A, Larmuseau MHD, et al. 2018. Determining Y-STR Mutation Rates in Deep-Routing Genealogies: Identification of Haplogroup Differences. *Forensic Science International: Genetics* 34(November 2017):1–10.

Čokić VP, Kecmanović M, Zgonjanin Bosić D, Jakovski Z, Veljković A, Katić S, et al. 2019. A Comprehensive Mutation Study in Wide Deep-Rooted R1b Serbian Pedigree: Mutation Rates and Male Relative Differentiation Capacity of 36 Y-STR Markers. *Forensic Science International: Genetics* 41(April):137–44.

Cox MP and Lahr MM. 2006. Y-Chromosome Diversity Is Inversely Associated with Language Affiliation in Paired Austronesian- and Papuan-Speaking Communities from Solomon Islands. *American Journal of Human Biology* 18(1):35–50.

Cristofaro J Di, Mazières S, Tous A, Gaetano C Di, Lin AA, Nebbia P, et al. 2018. Prehistoric Migrations through the Mediterranean Basin Shaped Corsican Y-Chromosome Diversity. *PLoS ONE* 13(8):1–15.

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, et al. 2004. Phylogeographic Analysis of Haplogroup E3b (E-M215) Y Chromosomes Reveals Multiple Migratory Events Within and Out of Africa. *American Journal of Human Genetics* 74(5):1014–22.

Cruciani F, La Fratta R, Trombetta B, Santolamazza P, Sellitto D, Colomb EB, et al. 2007. Tracing Past Human Male Movements in Northern/Eastern Africa and Western Eurasia: New Clues from Y-Chromosomal Haplogroups E-M78 and J-M12. *Molecular Biology and Evolution* 24(6):1300–1311.

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, et al. 2002. A Back Migration from Asia to Sub-Saharan Africa Is Supported by High-Resolution Analysis of Human Y-Chromosome Haplotypes. *American Journal of Human Genetics* 70(5):1197–1214.

Cruciani F, Trombetta B, Antonelli C, Pascone R, Valesini G, Scalzi V, et al. 2011. Strong Intra- and Inter-Continental Differentiation Revealed by Y Chromosome SNPs M269, U106 and U152. *Forensic Science International: Genetics* 5(3):e49–52.

Cruciani F, Trombetta B, Macaulay V, and Scozzari R. 2010. About the X-to-Y Gene Conversion Rate. *American Journal of Human Genetics* 86(3):495–97.

Cruciani F, Trombetta B, Massaia A, Destro-Bisol G, Sellitto D, and Scozzari R. 2011. A Revised Root for the Human Y Chromosomal Phylogenetic Tree: The Origin of Patrilineal Diversity in Africa. *American Journal of Human Genetics* 88(6):814–18.

Csáky V, Gerber D, Koncz I, Csiky G, Mende BG, Szeifert B, et al. 2020. Genetic Insights into the Social Organisation of the Avar Period Elite in the 7th Century AD Carpathian Basin. *Scfientific Reports* 10(948):415760.

Csányi B, Bogácsi-Szabo E, Tömöry G, Czibula Á, Priskin K, Csõsz A, et al. 2008. Y-Chromosome Analysis of Ancient Hungarian and Two Modern Hungarian-Speaking Populations from the Carpathian Basin. *Annals of Human Genetics* 72(4):519–34.

Cui Y, Li H, Ning C, Zhang Y, Chen L, Zhao X, et al. 2013. Y Chromosome Analysis of Prehistoric Human Populations in the West Liao River Valley, Northeast China. *BMC Evolutionary Biology* 13(1).

Delfin F, Myles S, Choi Y, Hughes D, Illek R, Van Oven M, et al. 2012. Bridging Near and Remote Oceania: MtDNA and NRY Variation in the Solomon Islands. *Molecular Biology and Evolution* 29(2):545–64.

Derenko M, Malyarchuk B, Denisova G, Wozniak M, Dambueva I, Dorzhu C, et al. 2006. Contrasting Patterns of Y-Chromosome Variation in South Siberian Populations from Baikal and Altai-Sayan Regions. *Human Genetics* 118(5):591–604.

Derenko M, Malyarchuk B, Denisova G, Wozniak M, Grzybowski T, Dambueva I, et al. 2007. Y-Chromosome Haplogroup N Dispersals from South Siberia to Europe. *Journal of Human Genetics* 52(9):763–70.

Dulik MC, Osipova LP, and Schurr TG. 2011. Y-Chromosome Variation in Altaian Kazakhs Reveals a Common Paternal Gene Pool for Kazakhs and the Influence of Mongolian Expansions. *PLoS ONE* 6(3).

Dulik MC, Zhadanov SI, Osipova LP, Askapuli A, Gau L, Gokcumen O, et al. 2012. Mitochondrial DNA and Y Chromosome Variation Provides Evidence for a Recent Common Ancestry between Native Americans and Indigenous Altaians. *American Journal of Human Genetics* 90(2):229–46.

Egyed B, Brandstätter A, Irwin JAJA, Pádár Z, Parsons TJT., Parson W, et al. 2007. Mitochondrial Control Region Sequence Variations in the Hungarian Population: Analysis of Population Samples from Hungary and from Transylvania (Romania). *Forensic Science International: Genetics* 1(2):158–62.

Fechner A, Quinque D, Rychkov S, Morozowa I, Naumova O, Schneider Y, et al. 2008. Boundaries and Clines in the West Eurasian Y-Chromosome Landscape: Insights from the European Part of Russia. *American Journal of Physical Anthropology* 137(1):41–47.

Fedorova SA, Reidla M, Metspalu E, Metspalu M, Rootsi S, Tambets K, et al. 2013. Autosomal and Uniparental Portraits of the Native Populations of Sakha (Yakutia): Implications for the Peopling of Northeast Eurasia. *BMC Evolutionary Biology* 13(1):1.

Fernandes DM, Strapagiel D, Borówka P, Marciniak B, Żądzińska E, Sirak K, et al. 2018. A Genomic Neolithic Time Transect of Hunter-Farmer Admixture in Central Poland. *Scientific Reports* 8(1):1–11.

Filatov DA, Monéger F, Negrutlu I, and Charlesworth D. 2000. Low Variability in a Y-Linked Plant Gene and Its Implications for Y- Chromosome Evolution. *Nature* 404(6776):388–90.

Finocchio A, Trombetta B, Messina F, D'Atanasio E, Akar N, Loutradis A, et al. 2018. A Finely Resolved Phylogeny of y Chromosome Hg J Illuminates the Processes of Phoenician and Greek Colonizations in the Mediterranean. *Scientific Reports* 8(1):3–11.

Fodor I. 1977. Où Le Dominicain Julien de Hongrie Retrouva-t-Il Les Hongrois de l'Est? Pp. 9–20 in *Les Anciens Hongrois Et Les Ethnies Voisines À L'Est*. Budapest: Akadémiai Kiadó

Fodor I. 2015. *Венгры: Древняя История И Обретение Родины*. Пермь: Зебра

Fóthi E. 2000. Anthropological Conclusions of the Study of Roman and Migration Periods. *Acta Biologica Szegediensis* 44(1–4):87–94.

Fóthi E, Fehér T, Fóthi Á, and Keyser C. 2019. *Európai És Ázsiai Apai Genetikai Vonalak A Honfoglaló Magyar Törzsekben*. Piliscsaba: Avicenna Institute of Middle Eastern Studies

Fóthi E, Gonzalez A, Fehér T, Gugora A, Fóthi Á, Biró O, et al. 2020. Genetic Analysis of Male Hungarian Conquerors: European and Asian Paternal Lineages of the Conquering Hungarian Tribes. *Archaeological and Anthropological Sciences* 12(1).

Francalacci P, Morelli L, Angius A, Berutti R, Reinier F, Atzeni R, et al. 2013. Low-Pass DNA Sequencing of 1200 Sardinians Reconstructs European Y-Chromosome Phylogeny. *Science* 341(August):565–69.

Francalacci P, Sanna D, and Useli A. 2016. Human Y Chromosome Mutation Rate: Problems and Perspectives. *Anthropology: Current and Future Developments* 2:65–91.

Da Fré NN, Rodenbusch R, Gastaldo AZ, Hanson E, Ballantyne J, and Alho CS. 2015. Genetic Data and de Novo Mutation Rates in Father-Son Pairs of 23 Y-STR Loci in Southern Brazil Population. *International Journal of Legal Medicine* 129(6): 1221–23.

Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, et al. 2015. An Early Modern Human from Romania with a Recent Neanderthal Ancestor. *Nature* 524(7564):216–19.

Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, et al. 2014. Genome Sequence of a 45,000-Year-Old Modern Human from Western Siberia. *Nature* 514(7253):445–49.

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, et al. 2016. The Genetic History of Ice Age Europe. *Nature* 534(7606):200–205.

Galushkin SK, Spitsyn VA, and Crawford MH. 2001. Genetic Structure of Mongolic-Speaking Kalmyks. *Human Biology* 73(6):823–34.

Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, et al. 2014. Genome Flux and Stasis in a Five Millennium Transect of European Prehistory. *Nature Communications* 5:1–9.

Garustovic GN. 1988. Об Этнической Принадлежности Раннемусульманских Памятников Западной и Центральной Башкирии. P. 130 in *Проблемы Древних Угров На Южном Урале*. Ufa: БНЦ УрО АН СССР

Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovsky LA, Underhill PA, et al. 2007. The Himalayas as a Directional Barrier to Gene Flow. *American Journal of Human Genetics* 80(5):884–94.

Ge J, Budowle B, Aranda XG, Planz J V., Eisenberg AJ, and Chakraborty R. 2009. Mutation Rates at Y Chromosome Short Tandem Repeats in Texas Populations. *Forensic Science International: Genetics* 3(3):179–84.

Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, et al. 2004. Y Chromosomal Haplogroup J as a Signature of the Post-Neolithic Colonization of Europe. *Human Genetics* 115(5):357–71.

Goedbloed M, Vermeulen M, Fang RN, Lembring M, Wollstein A, Ballantyne K, et al. 2009. Comprehensive Mutation Analysis of 17 Y-Chromosomal Short Tandem Repeat Polymorphisms Included in the AmpFSTR® Yfiler® PCR Amplification Kit. *International Journal of Legal Medicine* 123(6):471–82.

Gonçalves R, Freitas A, Branco M, Rosa A, Fernandes AT, and Brehm A. 2006. Y-Chromosome Lineages from Portugal, Madeira and Açores Record Elements of Sephardim and Berber Ancestry. *International Congress Series* 1288:237–39.

González-Fortes G, Jones ER, Lightfoot E, Bonsall C, Lazar C, Grandal-d'Anglade A, et al. 2017. Paleogenomic Evidence for Multi-Generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin. *Current Biology* 27(12):1801-1810.e10.

Goodwin S, McPherson JD, and McCombie WR. 2016. Coming of Age: Ten Years of next-Generation Sequencing Technologies. *Nature Reviews Genetics* 17(6): 333–51.

Graves M. 1995. The Origin and Function of the Mammalian Y Chromosome and Y-borne Genes – an Evolving Understanding. *BioEssays* 17(4):311–20.

Grugni V, Raveane A, Ongaro L, Battaglia V, Trombetta B, Colombo G, et al. 2019. Analysis of the Human Y-Chromosome Haplogroup Q Characterizes Ancient Population Movements in Eurasia and the Americas. *BMC Biology* 17(1):1–14.

Guillot EG and Cox MP. 2015. High Frequency Haplotypes Are Expected Events, Not Historical Figures. *F1000Research* 4(666):1–9.

Günther T, Malmström H, Svensson EM, Omrak A, Sánchez-Quinto F, Kılınç GM, et al. 2018. Population Genomics of Mesolithic Scandinavia: Investigating Early Postglacial Migration Routes and High-Latitude Adaptation. *PLoS Biology* 16(1): 1–22.

Günther T, Valdiosera C, Malmström H, Ureña I, Rodriguez-Varela R, Sverrisdóttir ÓO, et al. 2015. Ancient Genomes Link Early Farmers from Atapuerca in Spain to Modern-Day Basques. *Proceedings of the National Academy of Sciences of the United States of America* 112(38):11917–22.

Gusmão A, Gusmão L, Gomes V, Alves C, Calafell F, Amorim A, et al. 2008. A Perspective on the History of the Iberian Gypsies Provided by Phylogeographic Analysis of Y-Chromosome Lineages. *Annals of Human Genetics* 72(2):215–27.

Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, et al. 2010. Ancient DNA from European Early Neolithic Farmers Reveals Their near Eastern Affinities. *PLoS Biology* 8(11).

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. 2015. Massive Migration from the Steppe Was a Source for Indo-European Languages in Europe. *Nature* 522(7555):207–11.

Haber M, Jones AL, Connell BA, Asan, Arciero E, Yang H, et al. 2019. A Rare Deep-Rooting D0 African Y-Chromosomal Haplogroup and Its Implications for the Expansion of Modern Humans out of Africa. *Genetics* 212(4):1421–28.

Hallast P, Batini C, Zadik D, Delser PM, Wetton JH, Arroyo-Pardo E, et al. 2014. The Y-Chromosome Tree Bursts into Leaf: 13,000 High-Confidence SNPs Covering the Majority of Known Clades. *Molecular Biology and Evolution* 32(3):661–73.

Hammer MF. 1995. A Recent Common Ancestry for Human Y Chromosomes. *Nature* 378(6555):376–78.

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, et al. 1998. Out of Africa and Back Again: Nested Cladistic Analysis of Human Y Chromosome Variation. *Molecular Biology and Evolution* 15(4):427–41.

Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, et al. 2006. Dual Origins of the Japanese: Common Ground for Hunter-Gatherer and Farmer Y Chromosomes. *Journal of Human Genetics* 51(1):47–58.

Harley VR, Jackson DI, Hextall PJ, Hawkins JR, Berkovitz GD, Sockanathan S, et al. 1992. DNA Binding Activity of Recombinant SRY from Normal Males and XY Females. *Science* 255(5043):453–56.

Helgason A, Einarsson AW, Gumundsdóttir VB, Sigursson Á, Gunnarsdóttir ED, Jagadeesan A, et al. 2015. The Y-Chromosome Point Mutation Rate in Humans. *Nature Genetics* 47(5):453–57.

Hernández CL, Dugoujon JM, Sánchez-Martínez LJ, Cuesta P, Novelletto A, and Calderón R. 2019. Paternal Lineages in Southern Iberia Provide Time Frames for Gene Flow from Mainland Europe and the Mediterranean World. *Annals of Human Biology* 46(1):63–76.

Heyer E, Puymirat J, Dieltjes P, Bakker E, and De Knijff P. 1997. Estimating Y Chromosome Specific Microsatellite Mutation Frequencies Using Deep Rooting Pedigrees. *Human Molecular Genetics* 6(5):799–803.

Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Díez-Del-Molino D, et al. 2016. Early Farmers from across Europe Directly Descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences of the United States of America* 113(25):6886–91.

Hohoff C, Dewa K, Sibbing U, Hoppe K, Forster P, and Brinkmann B. 2007. Y-Chromosomal Microsatellite Mutation Rates in a Population Sample from North-western Germany. *International Journal of Legal Medicine* 121(5):359–63.

Hovhannisyan A, Khachatryan Z, Haber M, Hrechdakian P, Karafet T, Zalloua P, et al. 2014. Different Waves and Directions of Neolithic Migrations in the Armenian Highland. *Investigative Genetics* 5(1):1–11.

Hu K, Yan S, Liu K, Ning C, Wei L, Li S, et al. 2015. The Dichotomy Structure of Y Chromosome Haplogroup N. *ArXiv*.

Huang Y-Z, Pamjav H, Flegontov P, Stenzl V, Wen SQ, Tong XZ, et al. 2018. Dispersals of the Siberian Y-Chromosome Haplogroup Q in Eurasia. *Molecular Genetics and Genomics* 293(1):107–17.

Huang Y-Z, Wei L-H, Yan S, Wen S-QQ, Wang C-CC, Yang Y-JJ, et al. 2018. Whole Sequence Analysis Indicates a Recent Southern Origin of Mongolian Y-Chromosome C2c1a1a1-M407. *Molecular Genetics and Genomics* 293(3):657–63.

Hudjashov G, Endicott P, Post H, Nagle N, Ho SYW, Lawson DJ, et al. 2018. Investigating the Origins of Eastern Polynesians Using Genome-Wide Data from the Leeward Society Isles. *Scientific Reports* 8(1).

Hughes JF and Page DC. 2015. The Biology and Evolution of Mammalian Y Chromosomes. *Annual Review of Genetics* 49(1):507–27.

Hughes JF and Rozen S. 2012. Genomics and Genetics of Human and Primate Y Chromosomes. *Annual Review of Genomics and Human Genetics* 13(1):83–108.

Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, Fulton RS, et al. 2012. Strict Evolutionary Conservation Followed Rapid Gene Loss on Human and Rhesus Y Chromosomes. *Nature* 483(7387):82–87.

Hughes JF, Skaletsky H, Pyntikova T, Graves TA, Van Daalen SKM, Minx PJ, et al. 2010. Chimpanzee and Human y Chromosomes Are Remarkably Divergent in Structure and Gene Content. *Nature* 463(7280):536–39.

Hughes JF, Skaletsky H, Pyntikova T, Minx PJ, Graves T, Rozen S, et al. 2005. Conservation of Y-Linked Genes during Human Evolution Revealed by Comparative Sequencing in Chimpanzee. *Nature* 437(7055):101–4.

Ilumäe AM, Reidla M, Chukhryaeva M, Järve M, Post H, Karmin M, et al. 2016. Human Y Chromosome Haplogroup N: A Non-Trivial Time-Resolved Phylogeography That Cuts across Language Families. *American Journal of Human Genetics* 99(1):163–73.

Inagaki S, Yamamoto Y, Doi Y, Takata T, Ishikawa T, Yoshitome K, et al. 2002. Typing of Y Chromosome Single Nucleotide Polymorphisms in a Japanese Population by a Multiplexed Single Nucleotide Primer Extension Reaction. *Legal Medicine* 4(3):202–6.

Ivanov VA, Belavin AM, and Krylasova NB. 2009. *Угры Предуралья В Древности И Средние Века*. Ufa: БГПУ

Janhunen J. 2006. The Mongolic Languages. Pp. 231–34 in *The encyclopedia of language & linguistics*. Amsterdam, Netherlands: Elsevier Scientific Publ. Co

Janhunen J. 2009. Proto-Uralic — What, Where, and When? Pp. 57–78 in *The Quasquicentennial of the Finno-Ugrian Society. Suomalais-Ugrilaisen Seuran Toimituksia. Mémoires de la Société Finno-Ougrienne*, edited by J. Ylikoski. Tiedekirja

Järve M, Saag L, Scheib CL, Pathak AK, Montinaro F, Pagani L, et al. 2019. Shifts in the Genetic Landscape of the Western Eurasian Steppe Associated with the Beginning and End of the Scythian Dominance. *Current Biology* 29(14):2430-2441.e10.

Jobling MA and Tyler-Smith C. 2003. The Human Y Chromosome: An Evolutionary Marker Comes of Age. *Nature Reviews Genetics* 4(8):598–612.

Jobling MA and Tyler-Smith C. 2017. Human Y-Chromosome Variation in the Genome-Sequencing Era. *Nature Reviews Genetics* 18(8):485–97.

Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, et al. 2015. Upper Palaeolithic Genomes Reveal Deep Roots of Modern Eurasians. *Nature Communications* 6:1–8.

Jones ER, Zarina G, Moiseyev V, Lightfoot E, Nigst PR, Manica A, et al. 2017. The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. *Current Biology* 27(4):576–82.

Juras A, Krzewińska M, Nikitin AG, Ehler E, Chyleński M, Łukasik S, et al. 2017. Diverse Origin of Mitochondrial Lineages in Iron Age Black Sea Scythians. *Scientific Reports* 7(March):1–10.

Juvaini A-M. 1997. *Genghis Khan: The History of the World Conqueror*. Manchester, UK: Manchester University Press

Karafet TM, Lansing JS, Redd AJ, Reznikova S, Watkins JC, Surata SPK, et al. 2005. Balinese Y-Chromosome Perspective on the Peopling of Indonesia: Genetic Contributions from Pre-Neolithic Hunter-Gatherers, Austronesian Farmers, and Indian Traders. *Human Biology* 77(1):93–114.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, and Hammer MF. 2008. New Binary Polymorphisms Reshape and Increase Resolution of the Human Y Chromosomal Haplogroup Tree. *Genome Research* 18(5):830–38.

Karafet TM, Osipova LP, Gubina MA, Posukh OL, Zegura SL, and 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. *Human Biology* 74(6):761–89.

Karafet TM, Osipova LP, Savina O V., Hallmark B, and Hammer MF. 2018. Siberian Genetic Diversity Reveals Complex Origins of the Samoyedic-Speaking Populations. *American Journal of Human Biology* 30(6):1–14.

Karlsson AO, Wallerström T, Götherström A, and Holmlund G. 2006. Y-Chromosome Diversity in Sweden – A Long-Time Perspective. *European Journal of Human Genetics* 14(8):963–70.

Karmin M, Saag L, Vicente M, Wilson Sayres MA, Järve M, Gerst Talas U, et al. 2015. A Recent Bottleneck of Y Chromosome Diversity Coincides with a Global Change in Culture. *Genome Res* 25:459–66.

Katoh T, Munkhbat B, Tounai K, Mano S, Ando H, Oyungerel G, et al. 2005. Genetic Features of Mongolian Ethnic Groups Revealed by Y-Chromosomal Analysis. *Gene* 346:63–70.

Katzner K. 1986. *The Languages of the World*. London, UK: Routledge and Kegan Paul

Kayser M, Brauer S, Cordaux R, Casto A, Lao O, Zhivotovsky LA, et al. 2006. Melanesian and Asian Origins of Polynesians: MtDNA and Y Chromosome Gradients across the Pacific. *Molecular Biology and Evolution* 23(11):2234–44.

Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill P, Shen P, et al. 2003. Reduced Y-Chromosome, but Not Mitochondrial DNA, Diversity in Human Populations from West New Guinea. *American Journal of Human Genetics* 72(2):281–302.

Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edelmann J, et al. 2005. Significant Genetic Differentiation between Poland and Germany Follows Present-Day Political Borders, as Revealed by Y-Chromosome Analysis. *Human Genetics* 117(5):428–43.

Kazakov EP. 1981. О Кушнаренковских Памятниках На Территории Татарии. in *Вопросы Археологии Урала*. Sverdlovsk (Jekaterinburg)

Kazakov EP. 1986. О Происхождении и Культурной Принадлежности Памятников с Гребенчато-Шнуровой Керамикой. in *Проблемы Средневековой Археологии Урала И Поволжья*. Ufa: БФАН СССР

Keller A, Graefen A, Ball M, Matzas M, Boisguerin V, Maixner F, et al. 2012. New Insights into the Tyrolean Iceman's Origin and Phenotype as Inferred by Whole-Genome Sequencing. *Nature Communications* 3.

Keža S. 1999. *Simon of Kéza, The Deeds of Hungarians*. edited by L. Veszprémy, F. Schaer, and J. Szűcs. Budapest; New-York: Central European University Press,

Kharkov VN, Khamina KV, Medvedeva OF, Simonova KV, Khitrinskaya IY, and Stepanov VA. 2013. Gene-Pool Structure of Tuvinians Inferred from Y-Chromosome Marker Data [in Russian]. *Genetika* 49(12):1416–25.

Khomyakova IA and Balinova NV. 2017. Anthropological Features of Torghuts and Derbets of Kalmykia and Western Mongolia: A Comparative Analysis. Pp. 15–32 in *Bulletin of Moscow University*. Moscow, Russia

King RJ, Dicristofaro J, Kouvatsi A, Triantaphyllidis C, Scheidel W, Myres NM, et al. 2011. The Coming of the Greeks to Provence and Corsica: Y-Chromosome Models of Archaic Greek Colonization of the Western Mediterranean. *BMC Evolutionary Biology* 11(1):1–8.

Kivisild T. 2017. The Study of Human Y Chromosome Variation through Ancient DNA. *Human Genetics* 136(5):529–46.

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, et al. 2003. The Genetic Heritage of the Earliest Settlers Persists Both in Indian Tribal and Caste Populations. *American Journal of Human Genetics* 72(2):313–32.

Klarić IM, Salihović MP, Lauc LB, Zhivotovsky LA, Rootsi S, and Janićijevic B. 2009. Dissecting the Molecular Architecture and Origin of Bayash Romani Patrilineages: Genetic Influences from South-Asia and the Balkans. *American Journal of Physical Anthropology* 138(3):333–42.

Klíma L. 2016. *Jürkák, Tormák, Merják: Szemelvények A Finnugor Nyelvű Népek Történetének Korai Forrásaiból*. MTA Bölcsészettudományi Kutatóközpont

Knapp M, Clarke AC, Horsburgh KA, and Matisoo-Smith EA. 2012. Setting the Stage – Building and Working in an Ancient DNA Laboratory. *Annals of Anatomy* 194(1):3–6.

Koivulehto J. 2001. The Earliest Contacts between Indo-European and Uralic Speakers in the Light of Lexical Loans. Pp. 235–64 in *Early contacts between Uralic and Indo-European: linguistic and archaeological considerations*, edited by C. Carpelan, A. Parpola, and P. Koskikallio. Helsinki: Suomalais-Ugrilaisen Seuran Toimituksia. Mémoires de la Société Finno-Ougrienne

Krzewińska M, Kılınç GM, Juras A, Koptekin D, Chyleński M, Nikitin AG, et al. 2018. Ancient Genomes Suggest the Eastern Pontic-Caspian Steppe as the Source of Western Iron Age Nomads. *Science Advances* 4(10).

Kushniarevich A, Sivitskaya L, Danilenko N, Novogrodskii T, Tsybovsky I, Kiseleva A, et al. 2013. Uniparental Genetic Heritage of Belarusians: Encounter of Rare Middle Eastern Matrilineages with a Central European Mitochondrial DNA Pool. *PLoS ONE* 8(6):1–14.

Kushniarevich A, Utevska O, Chuhryaeva M, Agdzhoyan A, Dibirova K, Uktveryte I, et al. 2015. Genetic Heritage of the Balto-Slavic Speaking Populations: A Synthesis of Autosomal, Mitochondrial and Y-Chromosomal Data. *PLoS ONE* 10(9).

Lacan M, Keyser C, Ricaut FX, Brucato N, Tarrús J, Bosch A, et al. 2011. Ancient DNA Suggests the Leading Role Played by Men in the Neolithic Dissemination. *Proceedings of the National Academy of Sciences of the United States of America* 108(45):18255–59.

Lahn BT and Page DC. 1999. Four Evolutionary Strata on the Human X Chromosome. *Science* 286(5441):964–67.

Lahn BT, Pearson NM, and Jegalian K. 2001. The Human Y Chromosome, in the Light of Evolution. *Nature Reviews Genetics* 2(3):207–16.

Lamnidis TC, Majander K, Jeong C, Salmela E, Wessman A, Moiseyev V, et al. 2018. Ancient Fennoscandian Genomes Reveal Origin and Spread of Siberian Ancestry in Europe. *Nature Communications* 9(1):5018.

Lappalainen T, Koivumäki S, Salmela E, Huoponen K, Sistonen P, Savontaus ML, et al. 2006. Regional Differences among the Finns: A Y-Chromosomal Perspective. *Gene* 376(2):207–15.

Lappalainen T, Laitinen V, Salmela E, Andersen P, Huoponen K, Savontaus ML, et al. 2008. Migration Waves to the Baltic Sea Region. *Annals of Human Genetics* 72(3):337–48.

Larmuseau MHD, Vanderheyden N, Van Geystelen A, Van Oven M, Kayser M, and Decorte R. 2014. Increasing Phylogenetic Resolution Still Informative for y Chromosomal Studies on West-European Populations. *Forensic Science International: Genetics* 9(1):179–85.

Lazaridis I. 2018. The Evolutionary History of Human Populations in Europe. *Current Opinion in Genetics and Development* 53:21–27.

Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, et al. 2016. Genomic Insights into the Origin of Farming in the Ancient Near East. *Nature* 536(7617):419–24.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, et al. 2014. Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans. *Nature* 513(7518):409–13.

Lee EJ, Makarewicz C, Renneberg R, Harder M, Krause-Kyora B, Müller S, et al. 2012. Emerging Genetic Patterns of the European Neolithic: Perspectives from a Late Neolithic Bell Beaker Burial Site in Germany. *American Journal of Physical Anthropology* 148(4):571–79.

Lee YG, Choi SC, Kang Y, Kim KM, Kang CS, and Kim C. 2019. Constructing a Reference Genome in a Single Lab: The Possibility to Use Oxford Nanopore Technology. *Plants* 8(8):1–13.

Lell JT, Sukerink RI, Starikovskaya YB, Su B, Jin L, Schurr TG, et al. 2002. The Dual Origin and Siberian Affinites of Native American Y Chromosome. *American Journal of Human Genetics* 70:192–206.

Li H, Zhao X, Zhao Y, Li C, Si D, Zhou H, et al. 2011. Genetic Characteristics and Migration History of a Bronze Culture Population in the West Liao-River Valley Revealed by Ancient DNA. *Journal of Human Genetics* 56(12):815–22.

Li J, Zhang Y, Zhao Y, Chen Y, Ochir A, Sarenbilige, et al. 2018. The Genome of an Ancient Rouran Individual Reveals an Important Paternal Lineage in the Donghu Population. *American Journal of Physical Anthropology* 166(4):895–905.

Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, et al. 2014. Human Paternal and Maternal Demographic Histories: Insights from High-Resolution Y Chromosome and MtDNA Sequences. *Investigative Genetics* 5(1):1–17.

Lipson M, Szécsényi-Nagy A, Mallick S, Pósa A, Stégmár B, Keerl V, et al. 2017. Parallel Palaeogenomic Transects Reveal Complex Genetic History of Early European Farmers. *Nature* 551(7680):368–72.

Malyarchuk B, Derenko M, Denisova G, Khoyt S, Woźniak M, Grzybowski T, et al. 2013. Y-Chromosome Diversity in the Kalmyks at the Ethnical and Tribal Levels. *Journal of Human Genetics* 58(12):804–11.

Malyarchuk B, Derenko M, Denisova G, Maksimov A, Wozniak M, Grzybowski T, et al. 2011. Ancient Links between Siberians and Native Americans Revealed by Subtyping the Y Chromosome Haplogroup Q1a. *Journal of Human Genetics* 56(8):583–88.

Malyarchuk B, Derenko M, Denisova G, Wozniak M, Grzybowski T, Dambueva I, et al. 2010. Phylogeography of the Y-Chromosome Haplogroup C in Northern Eurasia. *Annals of Human Genetics* 74(6):539–46.

Malyarchuk BA, Derenko M, Denisova G, Woźniak M, Rogalla U, Dambueva I, et al. 2016. Y Chromosome Haplotype Diversity in Mongolic-Speaking Populations and Gene Conversion at the Duplicated STR DYS385a,b in Haplogroup C3-M407. *Journal of Human Genetics* 61(6):491–96.

Manco L, Albuquerque J, Sousa MF, Martiniano R, de Oliveira RC, Marques S, et al. 2018. The Eastern Side of the Westernmost Europeans: Insights from Subclades within Y-Chromosome Haplogroup J-M304. *American Journal of Human Biology* 30(2):1–8.

Mangs HA and Morris B. 2007. The Human Pseudoautosomal Region (PAR): Origin, Function and Future. *Current Genomics* 8(2):129–36.

Mantere T, Kersten S, and Hoischen A. 2019. Long-Read Sequencing Emerging in Medical Genetics. *Frontiers in Genetics* 10(MAY):1–14.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. 2005. Genome Sequencing in Microfabricated High-Density Picolitre Reactors. *Nature* 437(7057):376–80.

Martiniano R, Cassidy LM, Ó'Maoldúin R, McLaughlin R, Silva NM, Manco L, et al. 2017. The Population Genomics of Archaeological Transition in West Iberia: Investigation of Ancient Substructure Using Imputation and Haplotype-Based Methods. *PLoS Genetics* 13(7):1–24.

Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, Mallick S, et al. 2018. The Genomic History of Southeastern Europe. *Nature* 555(7695):197–203.

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, et al. 2015. Genome-Wide Patterns of Selection in 230 Ancient Eurasians. *Nature* 528(7583):499–503.

Matveeva GI. 2007. О Культурном и Хронологическом Соотношении Памятников Кушнаренковского и Караякуповского Типов. in *Средневековая Археология Евразийских Степей*. Kazan

Mendez FL, Krahn T, Schrack B, Krahn AM, Veeramah KR, Woerner AE, et al. 2013. An African American Paternal Lineage Adds an Extremely Ancient Root to the Human y Chromosome Phylogenetic Tree. *American Journal of Human Genetics* 92(3):454–59.

Mendez FL, Poznik GD, Castellano S, and Bustamante CD. 2016. The Divergence of Neandertal and Modern Human y Chromosomes. *American Journal of Human Genetics* 98(4):728–34.

Mirabal S, Regueiro M, Cadenas AM, Cavalli-Sforza LL, Underhill PA, Verbenko DA, et al. 2009. Y-Chromosome Distribution within the Geo-Linguistic Landscape of Northwestern Russia. *European Journal of Human Genetics* 17(10): 1260–73.

Mittnik A, Wang CC, Pfrengle S, Daubaras M, Zariņa G, Hallgren F, et al. 2018. The Genetic Prehistory of the Baltic Sea Region. *Nature Communications* 9(1):1–11.

Mueller JL, Skaletsky H, Brown LG, Zaghlul S, Rock S, Graves T, et al. 2013. Independent Specialization of the Human and Mouse X Chromosomes for the Male Germ Line. *Nature Genetics* 45(9):1083–87.

Mumm S, Molini B, Terrell J, Srivastava A, and Schlessinger D. 1997. Evolutionary Features of the 4-Mb Xq21.3 XY Homology Region Revealed by a Map at 60-Kb Resolution. *Genome Research* 7(4):307–14.

Myres NM, Rootsi S, Lin AA, Järve M, King RJ, Kutuev I, et al. 2011. A Major Y-Chromosome Haplogroup R1b Holocene Era Founder Effect in Central and Western Europe. *European Journal of Human Genetics* 19(1):95–101.

Nakano K, Shiroma A, Shimoji M, Tamotsu H, Ashimine N, Ohki S, et al. 2017. Advantages of Genome Sequencing by Long-Read Sequencer Using SMRT Technology in Medical Area. *Human Cell* 30(3):149–61.

Nasidze I, Quinque D, Dupanloup I, Cordaux R, Kokshunova L, and Stoneking M. 2005. Genetic Evidence for the Mongolian Ancestry of Kalmyks. *American Journal of Physical Anthropology* 128(4):846–54.

Nasidze I, Quinque D, Rahmani M, Alemohamad SA, Asadova P, Zhukova O, et al. 2009. MtDNA and Y-Chromosome Variation in the Talysh of Iran and Azerbaijan. *American Journal of Physical Anthropology* 138(1):82–89.

Nasidze I, Sarkisian T, Kerimov A, and Stoneking M. 2003. Testing Hypotheses of Language Replacement in the Caucasus: Evidence from the Y-Chromosome. *Human Genetics* 112(3):255–61.

Neparáczki E, Juhász Z, Pamjav H, Fehér T, Csányi B, Zink A, et al. 2017. Genetic Structure of the Early Hungarian Conquerors Inferred from MtDNA Haplotypes and Y-Chromosome Haplogroups in a Small Cemetery. *Molecular Genetics and Genomics* 292(1):201–14.

Neparáczki E, Maróti Z, Kalmár T, Maár K, Nagy I, Latinovics D, et al. 2019. Y-Chromosome Haplogroups from Hun, Avar and Conquering Hungarian Period Nomadic People of the Carpathian Basin. *Scientific Reports* 9(1):1–12.

Niederstätter H, Berger B, Erhart D, Willuweit S, Geppert M, Gassner C, et al. 2013. Multiple Recurrent Mutations at Four Human Y-Chromosomal Single Nucleotide Polymorphism Sites in a 37 Bp Sequence Tract on the ARSDP1 Pseudogene. *Forensic Science International: Genetics* 7(6):593–600.

Nielsen R and Beaumont MA. 2009. Statistical Inferences in Phylogeography. *Molecular Ecology* 18(6):1034–47.

Ning C, Li Tianjiao, Wang K, Zhang F, Li Tao, Wu Xiyan, et al. 2020. Ancient Genomes from Northern China Suggest. *Nature Communications* 11(2700).

O'Sullivan N, Posth C, Coia V, Schuenemann VJ, Douglas Price T, Wahl J, et al. 2018. Ancient Genome-Wide Analyses Infer Kinship Structure in an Early Medieval Alemannic Graveyard. *Science Advances* 4(9).

Ohno S. 1966. *Sex Chromosomes and Sex-Linked Genes*. Berlin: Springer, Berlin, Heidelberg

Olalde I, Allentoft ME, Sánchez-Quinto F, Santpere G, Chiang CWK, DeGiorgio M, et al. 2014. Derived Immune and Ancestral Pigmentation Alleles in a 7,000-Year-Old Mesolithic European. *Nature* 507(7491):225–28.

Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Booth T, et al. 2018. The Beaker Phenomenon and the Genomic Transformation of Northwest Europe. *Nature* 555(7695):190–96.

Omrak A, Günther T, Valdiosera C, Svensson EM, Malmström H, Kiesewetter H, et al. 2016. Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Current Biology* 26(2):270–75.

Orlando L, Gilbert MTP, and Willerslev E. 2015. Reconstructing Ancient Genomes and Epigenomes. *Nature Reviews Genetics* 16(7):395–408.

Otto SP, Pannell JR, Peichel CL, Ashman TL, Charlesworth D, Chippindale AK, et al. 2011. About PAR: The Distinct Evolutionary Dynamics of the Pseudoautosomal Region. *Trends in Genetics* 27(9):358–67.

van Oven M, Brauer S, Choi Y, Ensing J, Schiefenhövel W, Stoneking M, et al. 2014. Human Genetics of the Kula Ring: Y-Chromosome and Mitochondrial DNA Variation in the Massim of Papua New Guinea. *European Journal of Human Genetics : EJHG* 22(12):1393–1403.

Van Oven M, Van Geystelen A, Kayser M, Decorte R, and Larmuseau MH. 2014. Seeing the Wood for the Trees: A Minimal Reference Phylogeny for the Human Y Chromosome. *Human Mutation* 35(2):187–91.

Pääbo S. 1989. Ancient DNA: Extraction, Characterization, Molecular Cloning, and Enzymatic Amplification. *Proceedings of the National Academy of Sciences of the United States of America* 86(6):1939–43.

Pakendorf B, Novgorodov IN, Osakovskij VL, Danilova AP, Protod'jakonov AP, and Stoneking M. 2006. Investigating the Effects of Prehistoric Migrations in Siberia: Genetic Variation and the Origins of Yakuts. *Human Genetics* 120(3):334–53.

Pamjav H, Fehér T, Németh E, and Pádár Z. 2012. Brief Communication: New Y-Chromosome Binary Markers Improve Phylogenetic Resolution within Haplogroup R1a1. *American Journal of Physical Anthropology* 149(4):611–15.

Pamjav H, Fóthi, Fehér T, and Fóthi E. 2017. A Study of the Bodrogköz Population in North-Eastern Hungary by Y Chromosomal Haplotypes and Haplogroups. *Molecular Genetics and Genomics* 292(4):883–94.

Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Børresen-Dale AL, and Underhill PA. 2002. Different Genetic Components in the Norwegian Population Revealed by the Analysis of MtDNA and Y Chromosome Polymorphisms. *European Journal of Human Genetics* 10(9):521–29.

Pereira JB, Costa MD, Vieira D, Pala M, Bamford L, Harich N, et al. 2017. Reconciling Evidence from Ancient and Contemporary Genomes: A Major Source for the European Neolithic within Mediterranean Europe. *Proceedings of the Royal Society B: Biological Sciences* 284(1851).

Pfeifer SP. 2017. From Next-Generation Resequencing Reads to a High-Quality Variant Data Set. *Heredity* 118(2):111–24.

Pinotti T, Bergström A, Geppert M, Bawn M, Ohasi D, Shi W, et al. 2019. Y Chromosome Sequences Reveal a Short Beringian Standstill, Rapid Expansion, and Early Population Structure of Native American Founders. *Current Biology* 29(1):149-157.e3.

Poznik GD, Henn BM, Yee MC, Sliwerska E, Euskirchen GM, Lin AA, et al. 2013. Sequencing Y Chromosomes Resolves Discrepancy in Time to Common Ancestor of Males versus Females. *Science* 341(6145):562–65.

Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, et al. 2016. Punctuated Bursts in Human Male Demography Inferred from 1,244 Worldwide Y-Chromosome Sequences. *Nature Genetics* 48(6):593–99.

Pritchard JK, Seielstad MT, Perez-Lezaun A, and Feldman MW. 1999. Population Growth of Human Y Chromosomes: A Study of y Chromosome Microsatellites. *Molecular Biology and Evolution* 16(12):1791–98.

Racimo F, Woodbridge J, Fyfe RM, Sikora M, Sjögren K-G, Kristiansen K, et al. 2020. The Spatiotemporal Spread of Human Migrations during the European Holocene. *Proceedings of the National Academy of Sciences* 117(16):201920051.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, Degiorgio M, Stafford TW, et al. 2014. The Genome of a Late Pleistocene Human from a Clovis Burial Site in Western Montana. *Nature* 506(7487):225–29.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, et al. 2010. Ancient Human Genome Sequence of an Extinct Palaeo-Eskimo. *Nature* 463(7282): 757–62.

Rębała K, Martínez-Cruz B, Tönjes A, Kovacs P, Stumvoll M, Lindner I, et al. 2013. Contemporary Paternal Genetic Landscape of Polish and German Populations: From Early Medieval Slavic Expansion to Post-World War II Resettlements. *European Journal of Human Genetics* 21(4):415–22.

Rice WR. 1996. Evolution of the Y Sex Chromosome in Animals. *BioScience* 46(5): 331–43.

Roewer L, Krüger C, Willuweit S, Nagy M, Rodig H, Kokshunova L, et al. 2007. Y-Chromosomal STR Haplotypes in Kalmyk Population Samples. *Forensic Science International* 173(2–3):204–9.

Róna-Tas A. 1999. *Hungarians and Europe in the Early Middle Ages. An Introduction to Early Hungarian History*. Budapest: Central European University Press

Róna-Tas A and Árpád B. 2011. *Old Turkic. Turkic Loanwords In Hungarian*. 84th ed. Wiesbaden: Harrassowitz

Rootsi S, Behar DM, Järve M, Lin AA, Myres NM, Passarelli B, et al. 2013. Phylogenetic Applications of Whole Y-Chromosome Sequences and the Near Eastern Origin of Ashkenazi Levites. *Nature Communications* 4.

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, et al. 2004. Phylogeography of Y-Chromosome Haplogroup I Reveals Distinct Domains of Prehistoric Gene Flow in Europe. *American Journal of Human Genetics* 75:128–37.

Rootsi S, Myres NM, Lin AA, Järve M, King RJ, Kutuev I, et al. 2012. Distinguishing the Co-Ancestries of Haplogroup G Y-Chromosomes in the Populations of Europe and the Caucasus. *European Journal of Human Genetics* 20(12):1275–82.

Rootsi S, Zhivotovsky LA, Baldovič M, Kayser M, Kutuev IA, Khusainova R, et al. 2007. A Counter-Clockwise Northern Route of the Y-Chromosome Haplogroup N from Southeast Asia towards Europe. *European Journal of Human Genetics* 15(2):204–11.

Ross MT, Bentley DR, and Tyler-Smith C. 2006. The Sequences of the Human Sex Chromosomes. *Current Opinion in Genetics and Development* 16(3):213–18.

Rosser ZH, Balaresque P, and Jobling MA. 2009. Gene Conversion between the X Chromosome and the Male-Specific Region of the Y Chromosome at a Translocation Hotspot. *American Journal of Human Genetics* 85(1):130–34.

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, et al. 2000. Y-Chromosomal Diversity in Europe Is Clinal and Influenced Primarily by Geography, Rather than by Language. *The American Journal of Human Genetics* 67(6):1526–43.

Rothe J, Melisch C, Powers N, Geppert M, Zander J, Purps J, et al. 2015. Genetic Research at a Fivefold Children's Burial from Medieval Berlin. *Forensic Science International: Genetics* 15:90–97.

Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, et al. 2003. Abundant Gene Conversion between Arms of Palindromes in Human and Ape Y Chromosomes. *Nature* 423(6942):873–76.

Saag L, Laneman M, Varul L, Malve M, Valk H, Razzak MA, et al. 2019. The Arrival of Siberian Ancestry Connecting the Eastern Baltic to Uralic Speakers Further East. *Current Biology* 29(10):1701-1711.e16.

Saag L, Varul L, Scheib CL, Stenderup J, Allentoft ME, Saag LL, et al. 2017. Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe. *Current Biology* 27(14):2185-2193.e6.

Sánchez-Diz P, Alves C, Carvalho E, Carvalho M, Espinheira R, García O, et al. 2008. Population and Segregation Data on 17 Y-STRs: Results of a GEP-ISFG Collaborative Study. *International Journal of Legal Medicine* 122(6):529–33.

Sánchez-Quinto F, Malmström H, Fraser M, Girdland-Flink L, Svensson EM, Simões LG, et al. 2019. Megalithic Tombs in Western and Northern Neolithic Europe Were Linked to a Kindred Society. *Proceedings of the National Academy of Sciences of the United States of America* 116(19):9469–74.

Schwartz A, Chan DC, Brown LG, Alagappan R, Pettay D, Disteche C, et al. 1998. Reconstructing Hominid Y Evolution: X-Homologous Block, Created by X-Y Transposition, Was Disrupted by Yp Inversion through LINE-LINE Recombination. *Human Molecular Genetics* 7(1):1–11.

Scozzari R, Massaia A, Trombetta B, Bellusci G, Myres NM, Novelletto A, et al. 2014. An Unbiased Resource of Novel SNP Markers Provides a New Chronology for the Human y Chromosome and Reveals a Deep Phylogenetic Structure in Africa. *Genome Research* 24(3):535–44.

Sedov V V. 1987. Венгры в Восточной Европе. in *Финно-Угры И Балты В Эпоху Средневековья*. Moscow: Археология СССР

Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas AS, Manica A, Moltke I, et al. 2014. Genomic Structure in Europeans Dating Back at Least 36,200 Years Andaine. *Science* 346(6213):1113–18.

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, 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. *American Journal of Human Genetics* 74(5):1023–34.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, et al. 2000. The Genetic Legacy of Paleolithic Homo Sapiens Sapiens in Extant Europeans: A Y Chromosome Perspective. *Science* 290(5494):1155–59.

Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CET, et al. 2006. Polarity and Temporality of High-Resolution Y-Chromosome Distributions in India Identify Both Indigenous and Exogenous Expansions and Reveal Minor Genetic Influence of Central Asian Pastoralists. *American Journal of Human Genetics* 78(2):202–21.

Shi H, Dong YL, Wen B, Xiao CJ, Underbill PA, Shen PD, et al. 2005. Y-Chromosome Evidence of Southern Origin of the East Asian-Specific Haplogroup O3-M122. *American Journal of Human Genetics* 77(3):408–19.

Shi H, Qi X, Zhong H, Peng Y, Zhang X, Ma RZ, et al. 2013. Genetic Evidence of an East Asian Origin and Paleolithic Northward Migration of Y-Chromosome Haplogroup N. *PLoS ONE* 8(6):1–9.

Shi H, Zhong H, Peng Y, Dong YL, Qi X Bin, Zhang F, et al. 2008. Y Chromosome Evidence of Earliest Modern Human Settlement in East Asia and Multiple Origins of Tibetan and Japanese Populations. *BMC Biology* 6(Figure 1):1–10.

Shi W, Ayub Q, Vermeulen M, Shao RG, Zuniga S, Van Der Gaag K, et al. 2010. A Worldwide Survey of Human Male Demographic History Based on Y-SNP and

Y-STR Data from the HGDP-CEPH Populations. *Molecular Biology and Evolution* 27(2):385–93.
Sikora M, Carpenter ML, Moreno-Estrada A, Henn BM, Underhill PA, Sánchez-Quinto F, et al. 2014. Population Genomic Analysis of Ancient and Modern Genomes Yields New Insights into the Genetic Ancestry of the Tyrolean Iceman and the Genetic Structure of Europe. *PLoS Genetics* 10(5).
Sikora M, Seguin-Orlando A, Sousa VC, Albrechtsen A, Korneliussen T, Ko A, et al. 2017. Ancient Genomes Show Social and Reproductive Behavior of Early Upper Paleolithic Foragers. *Science* 358(6363):659–62.
Singh S, Singh Ashish, Rajkumar R, Kumar KS, Samy SK, Nizamuddin S, et al. 2016. Dissecting the Influence of Neolithic Demic Diffusion on Indian Y-Chromosome Pool through J2-M172 Haplogroup. *Scientific Reports* 6(January):1–10.
Skaletsky H, Kuroda-Kawaguchl T, Minx PJ, Cordum HS, Hlllier LD, Brown LG, et al. 2003. The Male-Specific Region of the Human Y Chromosome Is a Mosaic of Discrete Sequence Classes. *Nature* 423(6942):825–37.
Skoglund P, Malmström H, Raghavan M, Storå J, Hall P, Willerslev E, et al. 2012. Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science* 336(6080):466–69.
Stringer C. 2002. Modern Human Origins: Progress and Prospects. *Philosophical Transactions of the Royal Society B: Biological Sciences* 357(1420):563–79.
Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, et al. 1999. Y-Chromosome Evidence for a Northward Migration of Modern Humans into Eastern Asia during the Last Ice Age. *American Journal of Human Genetics* 65(6):1718–24.
Szécsényi-Nagy A, Brandt G, Haak W, Keerl V, Jakucs J, Möller-Rieker S, et al. 2015. Tracing the Genetic Origin of Europe's First Farmers Reveals Insights into Their Social Organization. *Proceedings of the Royal Society B: Biological Sciences* 282(1805).
Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogväli E-L, et al. 2004. The Western and Eastern Roots of the Saami – The Story of Genetic 'Outliers' Told by Mitochondrial DNA and Y Chromosomes. *The American Journal of Human Genetics* 74(4):661–82.
Tambets K, Yunusbayev B, Hudjashov G, Ilumäe AM, Rootsi S, Honkola T, et al. 2018. Genes Reveal Traces of Common Recent Demographic History for Most of the Uralic-Speaking Populations. *Genome Biology* 19(1):1–20.
Tang H, Siegmund DO, Shen P, Oefner PJ, and Feldman MW. 2002. Frequentist Estimation of Coalescence Times from Nucleotide Sequence Data Using a Tree-Based Partition. *Genetics* 161(1):447–59.
Teitz LS, Pyntikova T, Skaletsky H, and Page DC. 2018. Selection Has Countered High Mutability to Preserve the Ancestral Copy Number of Y Chromosome Amplicons in Diverse Human Lineages. *American Journal of Human Genetics* 103(2):261–75.
Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, et al. 2005. Reconstructing the Origin of Andaman Islanders. *Science* 308(5724):996.
The Y Chromosome Consortium. 2002. A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome Research* 12(2):339–48.
Thomson R, Pritchard JK, Shen P, Oefner PJ, and Feldman MW. 2000. Recent Common Ancestry of Human Y Chromosomes: Evidence from DNA Sequence Data. *Proceedings of the National Academy of Sciences of the United States of America* 97(13):7360–65.

Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, et al. 2001. A Physical Map of the Human Y Chromosome. *Nature* 409(6822): 943–45.

Trejaut JA, Poloni ES, Yen JC, Lai YH, Loo JH, Lee CL, et al. 2014. Taiwan Y-Chromosomal DNA Variation and Its Relationship with Island Southeast Asia. *BMC Genetics* 15.

Trinkaus E, Milota Ş, Rodrigo R, Mircea G, and Moldovan O. 2003. Early Modern Human Cranial Remains from the Peştera Cu Oase, Romania. *Journal of Human Evolution* 45(3):245–53.

Trombetta B and Cruciani F. 2017. Y Chromosome Palindromes and Gene Conversion. *Human Genetics* 136(5):605–19.

Trombetta B, Cruciani F, Underhill PA, Sellitto D, and Scozzari R. 2010. Footprints of X-to-Y Gene Conversion in Recent Human Evolution. *Molecular Biology and Evolution* 27(3):714–25.

Trombetta B, D'Atanasio E, and Cruciani F. 2017. Patterns of Inter-Chromosomal Gene Conversion on the Male-Specific Region of the Human Y Chromosome. *Frontiers in Genetics* 8(MAY):1–7.

Trombetta B, Fantini G, D'Atanasio E, Sellitto D, and Cruciani F. 2016. Evidence of Extensive Non-Allelic Gene Conversion among LTR Elements in the Human Genome. *Scientific Reports* 6:1–11.

Trombetta B, Sellitto D, Scozzari R, and Cruciani F. 2014. Inter- and Intraspecies Phylogenetic Analyses Reveal Extensive X-Y Gene Conversion in the Evolution of Gametologous Sequences of Human Sex Chromosomes. *Molecular Biology and Evolution* 31(8):2108–23.

Türk A. 2012. New Archaeological Research Design for Early Hungarian History. *Hungarian Archaeology E-Journal* Summer:1–6.

Türk A, Boldog Z, Petkes Z, and Sudár B. 2015. From the Ancient Homelands to the Carpathians – from the Finds to the Apparel. *Hungarian Archaeology E-Journal* (Autumn).

Turrina S, Atzei R, and De Leo D. 2006. Y-Chromosomal STR Haplotypes in a Northeast Italian Population Sample Using 17plex Loci PCR Assay. *International Journal of Legal Medicine* 120(1):56–59.

Turrina S, Caratti S, Ferrian M, and De Leo D. 2015. Haplotype Data and Mutation Rates for the 23 Y-STR Loci of PowerPlex® Y 23 System in a Northeast Italian Population Sample. *International Journal of Legal Medicine* 129(4):725–28.

Underhill PA, Jin L, Lin AA, Qasim Mehdi S, Jenkins T, Vollrath D, et al. 1997. Detection of Numerous Y Chromosome Biallelic Polymorphisms by Denaturing High-Performance Liquid Chromatography. *Genome Research* 7(10):996–1005.

Underhill PA and Kivisild T. 2007. Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations. *Annual Review of Genetics* 41(1):539–64.

Underhill PA, Myres NM, Rootsi S, Chow CT, Lin AA, Otillar RP, et al. 2007. New Phylogenetic Relationships for Y-Chromosome Haplogroup I: Reappraising Its Phylogeography and Prehistory. *Rethinking the Human Evolution* 33–42.

Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovsky LA, King RJ, et al. 2010. Separating the Post-Glacial Coancestry of European and Asian y Chromosomes within Haplogroup R1a. *European Journal of Human Genetics* 18(4):479–84.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazón Lahr M, Foley RA, et al. 2001. The Phylogeography of Y Chromosome Binary Haplotypes and the Origins of Modern Human Populations. *Annals of Human Genetics* (1):43–62.

Underhill PA, Poznik GD, Rootsi S, Järve M, Lin AA, Wang J, et al. 2015. The Phylogenetic and Geographic Structure of Y-Chromosome Haplogroup R1a. *European Journal of Human Genetics* 23(1):124–31.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, et al. 2000. Y Chromosome Sequence Variation and the History of Human Populations. *Nature Genetics* 26(3):358–61.

Völgyi A, Zalán A, Szvetnik E, and Pamjav H. 2009. Hungarian Population Data for 11 Y-STR and 49 Y-SNP Markers. *Forensic Science International: Genetics* 3(2):e27–28.

Voskarides K, Mazières S, Hadjipanagi D, Di Cristofaro J, Ignatiou A, Stefanou C, et al. 2016. Y-Chromosome Phylogeographic Analysis of the Greek-Cypriot Population Reveals Elements Consistent with Neolithic and Bronze Age Settlements. *Investigative Genetics* 7(1).

Wang CC and Li H. 2013. Inferring Human History in East Asia from Y Chromosomes. *Investigative Genetics* 4(1):1.

Wang CC and Li H. 2015. Evaluating the Y Chromosomal STR Dating in Deep-Rooting Pedigrees. *Investigative Genetics* 6(1):8–10.

Wang CC, Yan S, Qin ZD, Lu Y, Ding QL, Wei LH, et al. 2013. Late Neolithic Expansion of Ancient Chinese Revealed by Y Chromosome Haplogroup O3a1c-002611. *Journal of Systematics and Evolution* 51(3):280–86.

Watanabe Y, Naka I, Khor SS, Sawai H, Hitomi Y, Tokunaga K, et al. 2019. Analysis of Whole Y-Chromosome Sequences Reveals the Japanese Population History in the Jomon Period. *Scientific Reports* 9(1):2–9.

Weale ME, Shah T, Jones AL, Greenhalgh J, Wilson JF, Nymadawa P, et al. 2003. Rare Deep-Rooting Y Chromosome Lineages in Humans: Lessons for Phylogeography. *Genetics* 165(1):229–34.

Wei LH, Wang LX, Wen SQ, Yan S, Canada R, Gurianov V, et al. 2018. Paternal Origin of Paleo-Indians in Siberia: Insights from Y-Chromosome Sequences. *European Journal of Human Genetics* 26(11):1687–96.

Wei LH, Yan S, Lu Y, Wen S-QQ, Huang Y-ZZ, Wang L-XX, et al. 2018. Whole-Sequence Analysis Indicates That the y Chromosome C2∗-Star Cluster Traces Back to Ordinary Mongols, Rather than Genghis Khan. *European Journal of Human Genetics* 26(2):230–37.

Wei LH, Yan S, Teo YY, Huang YZ, Wang LX, Yu G, et al. 2017. Phylogeography of Y-Chromosome Haplogroup O3a2b2-N6 Reveals Patrilineal Traces of Austronesian Populations on the Eastern Coastal Regions of Asia. *PLoS ONE* 12(4):1–12.

Wei W, Ayub Q, Chen Y, McCarthy S, Hou Y, Carbone I, et al. 2013. A Calibrated Human Y-Chromosomal Phylogeny Based on Resequencing. *Genome Research* 23(2):388–95.

Wei W, Ayub Q, Xue Y, and Tyler-Smith C. 2013. A Comparison of Y-Chromosomal Lineage Dating Using Either Resequencing or Y-SNP plus Y-STR Genotyping. *Forensic Science International: Genetics* 7(6):568–72.

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, et al. 2001. The Eurasian Heartland: A Continental Perspective on Y-Chromosome

Diversity. *Proceedings of the National Academy of Sciences of the United States of America* 98(18):10244–49.
Willems T, Gymrek M, Highnam G, Mittelman D, and Erlich Y. 2014. The Landscape of Human STR Variation. *Genome Research* 24(11):1894–1904.
Wilson IJ and Balding DJ. 1998. Genealogical Inference from Microsatellite Data. *Genetics* 150(1):499–510.
Wong EHM, Khrunin A, Nichols L, Pushkarev D, Khokhrin D, Verbenko D, et al. 2017. Reconstructing Genetic History of Siberian and Northeastern European Populations. *Genome Research* 27(1).
Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, et al. 2005. Contrasting Patterns of Y Chromosome and MtDNA Variation in Africa: Evidence for Sex-Biased Demographic Processes. *European Journal of Human Genetics* 13(7):867–76.
Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, et al. 2006. Male Demography in East Asia: A North-South Contrast in Human Population Expansion Times. *Genetics* 172(4):2431–39.
Yamamoto T, Senda T, Horiba D, Sakuma M, Kawaguchi Y, and Kano Y. 2013. Y-Chromosome Lineage in Five Regional Mongolian Populations. *Forensic Science International: Genetics Supplement Series* 4(1):e260–61.
Yan S, Wang CC, Li H, Li SL, and Jin L. 2011. An Updated Tree of Y-Chromosome Haplogroup O and Revised Phylogenetic Positions of Mutations P164 and PK4. *European Journal of Human Genetics* 19(9):1013–15.
Yan S, Wang CC, Zheng HX, Wang W, Qin ZD, Wei LH, et al. 2014. Y Chromosomes of 40% Chinese Descend from Three Neolithic Super-Grandfathers. *PLoS ONE* 9(8).
Yang Y, Wang W, Cheng F, Chen M, Chen T, Zhao J, et al. 2018. Haplotypic Polymorphisms and Mutation Rate Estimates of 22 Y-Chromosome STRs in the Northern Chinese Han Father-Son Pairs. *Scientific Reports* 8(1):6–11.
Ye D, Zaidi AA, Tomaszkiewicz M, Anthony K, Liebowitz C, DeGiorgio M, et al. 2018. High Levels of Copy Number Variation of Ampliconic Genes across Major Human Y Haplogroups. *Genome Biology and Evolution* 10(5):1333–50.
Yunusbayev B, Metspalu M, Ja M, Kutuev I, Rootsi S, Metspalu E, et al. 2012. The Caucasus as an Asymmetric Semipermeable Barrier to Ancient Human Migrations Research Article. *Molecular Biology and Evolution* 29(1):359–65.
Zakharov IA. 2010. A Search for a 'Genghis Khan' Chromosome. *Russian Journal of Genetics* 46(9):1130–31.
Zalloua PA, Platt DE, El Sibai M, Khalife J, Makhoul N, Haber M, et al. 2008. Identifying Genetic Traces of Historical Expansions: Phoenician Footprints in the Mediterranean. *American Journal of Human Genetics* 83(5):633–42.
Zegura SL, Karafet TM, Zhivotovsky LA, and Hammer MF. 2004. High-Resolution SNPs and Microsatellite Haplotypes Point to a Single, Recent Entry of Native American Y Chromosomes into the Americas. *Molecular Biology and Evolution* 21(1):164–75.
Zeng TC, Aw AJ, and Feldman MW. 2018. Cultural Hitchhiking and Competition between Patrilineal Kin Groups Explain the Post-Neolithic Y-Chromosome Bottleneck. *Nature Communications* 9(1).
Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, et al. 1997. Genetic Relationships of Asians and Northern Europeans, Revealed by Y- Chromosomal DNA Analysis. *American Journal of Human Genetics* 60(5):1174–83.

Zerjal T, Xue Y, Bertorelle G, Wells RSS, Bao W, Zhu S, et al. 2003. The Genetic Legacy of the Mongols. *The American Journal of Human Genetics* 72(3):717–21.
Zhang Y, Wu X, Li J, Li H, Zhao Y, and Zhou H. 2018. The Y-Chromosome Haplogroup C3∗-F3918, Likely Attributed to the Mongol Empire, Can Be Traced to a 2500-Year-Old Nomadic Group. *Journal of Human Genetics* 63(2):231–38.
Zhivotovsky LA, Underhill PA, Cinnioğlu C, Kayser M, Morar B, Kivisild T, et al. 2004. The Effective Mutation Rate at Y Chromosome Short Tandem Repeats, with Application to Human Population-Divergence Time. *American Journal of Human Genetics* 74(1):50–61.
Zhivotovsky LA, Underhill PA, and Feldman MW. 2006. Difference between Evolutionarily Effective and Germ Line Mutation Rate Due to Stochastically Varying Haplogroup Size. *Molecular Biology and Evolution* 23(12):2268–70.
Zhong H, Shi H, Qi X Bin, Duan ZY, Tan PP, Jin L, et al. 2011. Extended y Chromosome Investigation Suggests Postglacial Migrations of Modern Humans into East Asia via the Northern Route. *Molecular Biology and Evolution* 28(1):717–27.
Zhong H, Shi H, Qi X Bin, Xiao CJ, Jin L, Ma RZ, et al. 2010. Global Distribution of Y-Chromosome Haplogroup C Reveals the Prehistoric Migration Routes of African Exodus and Early Settlement in East Asia. *Journal of Human Genetics* 55(7):428–35.

Web resources

Ensembl – https://www.ensembl.org/
Ethnologue Languages of the World – https://www.ethnologue.com
Federal State Statistic Service – https://www.gks.ru
ISOGG – https://isogg.org
YFull – https://www.yfull.com/tree/

ACKNOWLEDGEMENTS

When I started this PhD journey 6 years ago, I had little understanding of what would lie ahead. There were times I was overwhelmed and even frightened. Fortunately, I had two extremely experienced supervisors on this road with me who guided me through all the hardships of the scientific life. Siiri, you have been my mentor since the master's studies and I am glad it has continued, as you never left my side and helped me overcome many difficulties I faced along the way. Richard, it is undisputed that the scientific knowledge and skill you possess is invaluable, and I am grateful that you have helped me scrutinize and view every scientific claim and detail more critically.

I also want to thank all the co-authors of the articles this dissertation is based on. I am particularly grateful to Endre Németh and Natalia Balinova who allowed me to take part in their research.

One very important part is the stronghold of Estonian Biocentre and the remarkable colleagues whom I want to thank from the bottom of my heart. The amazing variety of scientific ideas and friendly personalities kept me in high spirits and if a problem arose there was always a shoulder to cry on and help on the way. I would like to thank Lauri, Rodrigo, Hovik, Mahyuk and Francesco who helped me with the kitchen side of things. I am especially grateful to Anne-Mai whose knowledge and eye to detail came to my aid on the last leg of this expedition, and Monika who could be considered my unofficial supervisor. A big thanks goes also to all the current and previous PhD room roommates I have not yet mentioned – Ajai, Tina, Stefania, Linda, Burak, Katri and Mathilde.

I am also delighted to have the most amazing friends one could hope for. I want to thank Kadi, Triin and Klaarika for always providing support and laughter when I needed it. I hope you know how important that has been during my studies. Also, I am grateful to my best friend from childhood – Eveli. Although, I do not see you as often as I would like, I can always depend on you.

Lastly, I would definitely want express thanks to my parents who have always been my strength and shelter. Aitäh, ema ja isa, et koguaeg toeks olete olnud ja aidanud mind raskustega toime tulla. Ilma teieta ei oleks doktoriõpingutest ja sellest tööst asja saanud.

PUBLICATIONS

CURRICULUM VITAE

Name: Helen Post
Date of birth: April 30, 1990
Contacts: University of Tartu, Institute of Genomics, Riia 23b–306b, 51010 Tartu, Estonia, +372 737 5050
E-mail: helen.post@ut.ee

Education:

2014–2020 Doctoral studies, University of Tartu, Faculty of Science and Technology, Institute of Molecular and Cell Biology, Chair of Evolutionary Biology

2012–2014 *Msc*, University of Tartu, Faculty of Science and Technology, Institute of Molecular and Cell Biology, Chair of Evolutionary Biology, supervisors Maere Reidla, Siiri Rootsi "The role of Y-chromosomal haplogoup N in the formation of genepool of Uralic-speaking populations"

2009–2012 *BSc,* University of Tartu, Faculty of Science and Technology, Institute of Molecular and Cell Biology, Chair of Evolutionary Biology, supervisor Maere Reidla "Genetic characterization of Canadian Native American population Dakelh using mitochondrial DNA"

Professional employment:

2018–… University of Tartu, Institute of Genomics, Specialist

2012 Pärnu Hospital, Molecular Diagnostics, Laboratory technician

Additional information:

2019 Popular science article in the journal Tutulus

2016–… Tartu University Hospital Children's Foundation, Volunteer

2013–2015 "GenDi" application for Android platform

2014 Instructor of Molecular Biology Workshop for high school students

Publications:

Balinova N, **Post H**, Kushniarevich A, Flores R, Karmin M, Sahakyan H, Reidla M, Metspalu E, Litvinov S, Dzhaubermezov M, Akhmetova V, Khusainova R, Endicott P, Khusnutdinova E, Orlova K, Bakaeva E, Khomyakova I, Spitsina N, Zinchenko R, Villems R, Rootsi S. (2019). **Y-chromosomal analysis of clan structure of Kalmyks the only European Mongol people and their relationship to Oirat-Mongols of Inner Asia**. *European Journal of Human Genetics* 27:1466–1474.

Post H, Németh E, Klima L, Flores R, Fehér T, Türk A, Székely G, Sahakyan H, Mondal M, Montinaro F, Karmin M, Saag L, Yunusbayev B, Khusnutdinova EK, Metspalu E, Villems R, Tambets K, Rootsi S. (2019).

Y-chromosomal connection between Hungarians and geographically distant populations of the Ural Mountain region and West Siberia. *Scientific Reports* 9: 7786.

Hudjashov G, Endicott P, **Post H**, Nagle N, Ho SYW, Lawson DJ, Reidla M, Karmin M, Rootsi S, Metspalu E, Saag L, Villems R, Cox MP, Mitchell RJ, Garcia-Bertrand RL, Metspalu M, Herrera RJ. (2018). **Investigating the origins of eastern Polynesians using genome-wide data from the Leeward Society Isles**. *Scientific Reports* 8(1):1823–1823.

Ilumäe AM, Reidla M, Chukhryaeva M, Järve M, **Post H**, Karmin M, Saag L, Agdzhoyan A, Kushniarevich A, Litvinov S, Ekomasova N, Tambets K, Metspalu E, Khusainova R, Yunusbayev B, Khusnutdinova EK, Osipova LP, Fedorova S, Utevska O, Koshel S, Rootsi S. (2016). **Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families**. *American Journal of Human Genetics* 163–173.

ELULOOKIRJELDUS

Nimi: Helen Post
Sünniaeg: 30. aprill 1990
Kontakt: Tartu Ülikool, Genoomika Instituut Riia 23b–306b, 51010 Tartu, Eesti, +372 737 5050
E-post: helen.post@gmail.com

Hariduskäik:

2014–2020 doktoriõpingud geenitehnoloogias, Tartu Ülikooli Loodus- ja täppisteaduste valdkond, Molekulaar- ja rakubioloogia instituut

2012–2014 geenitehnoloogia magister, Tartu Ülikool, Loodus- ja tehnoloogiateaduskond, Molekulaar- ja rakubioloogia instituut, Evolutsioonilise bioloogia õppetool. Juhendajad Maere Reidla, Siiri Rootsi „Y-kromosomaalse haplogrupi N osa uurali keeli kõnelevate populatsioonide geenitiigi kujunemisel“

2009–2012 geenitehnoloogia bakalaureus, Tartu Ülikool, Loodus- ja tehnoloogiateaduskond, Molekulaar- ja rakubioloogia instituut, Evolutsioonilise bioloogia õppetool. Juhendaja Maere Reidla „Kanada põliselanike dakelhide geneetiline iseloomustus mitokondriaalse DNA põhjal”

Töökogemus:

2018–… spetsialist, Tartu Ülikooli genoomika instituut
2012 molekulaardiagnostika laboritehnik, Pärnu Haigla,

Lisainfo:

2019 Populaarteaduslik artikkel ajakirjas Tutulus
2016–… vabatahtlik, Tartu Ülikooli Kliinikumi Lastefond
2013–2015 “GenDi” rakendus Android platvormile
2014 Gümnaasiumide geneetika õppepäevade juhendaja Tartus

Teaduspublikatsioonid:
Loetletud inglisekeelse CV rubriigis *Publications*

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand**. The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro*. Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous grassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan**. Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.
41. **Sulev Ingerpuu**. Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.

42. **Veljo Kisand**. Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa**. Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa**. Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik**. Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo**. Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo**. Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots**. Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero**. Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees**. Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks**. Cholecystokinin (CCK) – induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and serotonin. Tartu, 1999, 123 p.
52. **Ebe Sild**. Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva**. Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna**. Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro**. Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane**. Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm**. Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg**. Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild**. The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu**. Studies of the TOL plasmid transcription factor XylS. Tartu, 2000, 88 p.
61. **Dina Lepik**. Modulation of viral DNA replication by tumor suppressor protein p53. Tartu, 2000, 106 p.
62. **Kai Vellak**. Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu, 2000, 122 p.

63. **Jonne Kotta**. Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu, 2000, 160 p.
64. **Georg Martin**. Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000, 139 p.
65. **Silvia Sepp**. Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira**. On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000, 96 p.
67. **Priit Zingel**. The role of planktonic ciliates in lake ecosystems. Tartu, 2001, 111 p.
68. **Tiit Teder**. Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu, 2001, 122 p.
69. **Hannes Kollist**. Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu, 2001, 80 p.
70. **Reet Marits**. Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu, 2001, 112 p.
71. **Vallo Tilgar**. Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002, 126 p.
72. **Rita Hõrak**. Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002, 108 p.
73. **Liina Eek-Piirsoo**. The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002, 74 p.
74. **Krõõt Aasamaa**. Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002, 110 p.
75. **Nele Ingerpuu**. Bryophyte diversity and vascular plants. Tartu, 2002, 112 p.
76. **Neeme Tõnisson**. Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002, 124 p.
77. **Margus Pensa**. Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003, 110 p.
78. **Asko Lõhmus**. Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003, 168 p.
79. **Viljar Jaks**. p53 – a switch in cellular circuit. Tartu, 2003, 160 p.
80. **Jaana Männik**. Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003, 140 p.
81. **Marek Sammul**. Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003, 159 p
82. **Ivar Ilves**. Virus-cell interactions in the replication cycle of bovine papillomavirus type 1. Tartu, 2003, 89 p.
83. **Andres Männik**. Design and characterization of a novel vector system based on the stable replicator of bovine papillomavirus type 1. Tartu, 2003, 109 p.

84. **Ivika Ostonen**. Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003, 158 p.
85. **Gudrun Veldre**. Somatic status of 12–15-year-old Tartu schoolchildren. Tartu, 2003, 199 p.
86. **Ülo Väli**. The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina*: taxonomy, phylogeography and ecology. Tartu, 2004, 159 p.
87. **Aare Abroi**. The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004, 135 p.
88. **Tiina Kahre**. Cystic fibrosis in Estonia. Tartu, 2004, 116 p.
89. **Helen Orav-Kotta**. Habitat choice and feeding activity of benthic suspension feeders and mesograzers in the northern Baltic Sea. Tartu, 2004, 117 p.
90. **Maarja Öpik**. Diversity of arbuscular mycorrhizal fungi in the roots of perennial plants and their effect on plant performance. Tartu, 2004, 175 p.
91. **Kadri Tali**. Species structure of *Neotinea ustulata*. Tartu, 2004, 109 p.
92. **Kristiina Tambets**. Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Tartu, 2004, 163 p.
93. **Arvi Jõers**. Regulation of p53-dependent transcription. Tartu, 2004, 103 p.
94. **Lilian Kadaja**. Studies on modulation of the activity of tumor suppressor protein p53. Tartu, 2004, 103 p.
95. **Jaak Truu**. Oil shale industry wastewater: impact on river microbial community and possibilities for bioremediation. Tartu, 2004, 128 p.
96. **Maire Peters**. Natural horizontal transfer of the *pheBA* operon. Tartu, 2004, 105 p.
97. **Ülo Maiväli**. Studies on the structure-function relationship of the bacterial ribosome. Tartu, 2004, 130 p.
98. **Merit Otsus**. Plant community regeneration and species diversity in dry calcareous grasslands. Tartu, 2004, 103 p.
99. **Mikk Heidemaa**. Systematic studies on sawflies of the genera *Dolerus, Empria,* and *Caliroa* (Hymenoptera: Tenthredinidae). Tartu, 2004, 167 p.
100. **Ilmar Tõnno**. The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N_2 fixation in some Estonian lakes. Tartu, 2004, 111 p.
101. **Lauri Saks**. Immune function, parasites, and carotenoid-based ornaments in greenfinches. Tartu, 2004, 144 p.
102. **Siiri Rootsi**. Human Y-chromosomal variation in European populations. Tartu, 2004, 142 p.
103. **Eve Vedler**. Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.
104. **Andres Tover**. Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 126 p.
105. **Helen Udras**. Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005, 100 p.

106. **Ave Suija**. Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005, 162 p.
107. **Piret Lõhmus**. Forest lichens and their substrata in Estonia. Tartu, 2005, 162 p.
108. **Inga Lips**. Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005, 156 p.
109. **Krista Kaasik**. Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005, 121 p.
110. **Juhan Javoiš**. The effects of experience on host acceptance in ovipositing moths. Tartu, 2005, 112 p.
111. **Tiina Sedman**. Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005, 103 p.
112. **Ruth Aguraiuja**. Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005, 112 p.
113. **Riho Teras**. Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 106 p.
114. **Mait Metspalu**. Through the course of prehistory in India: tracing the mtDNA trail. Tartu, 2005, 138 p.
115. **Elin Lõhmussaar**. The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006, 124 p.
116. **Priit Kupper**. Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006, 126 p.
117. **Heili Ilves**. Stress-induced transposition of Tn*4652* in *Pseudomonas Putida.* Tartu, 2006, 120 p.
118. **Silja Kuusk**. Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006, 126 p.
119. **Kersti Püssa**. Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006, 90 p.
120. **Lea Tummeleht**. Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006, 94 p.
121. **Toomas Esperk**. Larval instar as a key element of insect growth schedules. Tartu, 2006, 186 p.
122. **Harri Valdmann**. Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.
123. **Priit Jõers**. Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli**. Gata3 and Gata2 in inner ear development. Tartu, 2007, 123 p.
125. **Kai Rünk**. Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007, 143 p.

126. **Aveliina Helm**. Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007, 89 p.
127. **Leho Tedersoo**. Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007, 233 p.
128. **Marko Mägi**. The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007, 135 p.
129. **Valeria Lulla**. Replication strategies and applications of Semliki Forest virus. Tartu, 2007, 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007, 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007, 171 p.
132. **Tatjana Krama**. Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007, 112 p.
133. **Signe Saumaa**. The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida.* Tartu, 2007, 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007, 96 p.
135. **Priit Kilgas**. Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007, 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007, 95 p.
137. **Kärt Padari**. Protein transduction mechanisms of transportans. Tartu, 2008, 128 p.
138. **Siiri-Lii Sandre**. Selective forces on larval colouration in a moth. Tartu, 2008, 125 p.
139. **Ülle Jõgar**. Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008, 99 p.
140. **Lauri Laanisto**. Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008, 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008, 105 p.
142. **Birgot Paavel**. Bio-optical properties of turbid lakes. Tartu, 2008, 175 p.
143. **Kaire Torn**. Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg**. Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd**. Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.
146. **Lauri Saag**. Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.

147. **Ulvi Karu**. Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm**. Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks**. Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu**. Acclimation of stomatal structure and function in tree canopy: effect of light and CO_2 concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš**. Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko**. Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast**. Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats**. Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova**. The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida.* Tartu, 2009, 124 p.
157. **Tsipe Aavik**. Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2009, 112 p.
158. **Kaja Kiiver**. Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja**. Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.
160. **Pille Hallast**. Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB/CGB*) gene clusters: diversity and divergence of young duplicated genes. Tartu, 2009, 168 p.
161. **Ain Vellak**. Spatial and temporal aspects of plant species conservation. Tartu, 2009, 86 p.
162. **Triinu Remmel**. Body size evolution in insects with different colouration strategies: the role of predation risk. Tartu, 2009, 168 p.
163. **Jaana Salujõe**. Zooplankton as the indicator of ecological quality and fish predation in lake ecosystems. Tartu, 2009, 129 p.
164. **Ele Vahtmäe**. Mapping benthic habitat with remote sensing in optically complex coastal environments. Tartu, 2009, 109 p.
165. **Liisa Metsamaa**. Model-based assessment to improve the use of remote sensing in recognition and quantitative mapping of cyanobacteria. Tartu, 2009, 114 p.

166. **Pille Säälik**. The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
167. **Lauri Peil**. Ribosome assembly factors in *Escherichia coli*. Tartu, 2009, 147 p.
168. **Lea Hallik**. Generality and specificity in light harvesting, carbon gain capacity and shade tolerance among plant functional groups. Tartu, 2009, 99 p.
169. **Mariliis Tark**. Mutagenic potential of DNA damage repair and tolerance mechanisms under starvation stress. Tartu, 2009, 191 p.
170. **Riinu Rannap**. Impacts of habitat loss and restoration on amphibian populations. Tartu, 2009, 117 p.
171. **Maarja Adojaan**. Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.
172. **Signe Altmäe**. Genomics and transcriptomics of human induced ovarian folliculogenesis. Tartu, 2010, 179 p.
173. **Triin Suvi**. Mycorrhizal fungi of native and introduced trees in the Seychelles Islands. Tartu, 2010, 107 p.
174. **Velda Lauringson**. Role of suspension feeding in a brackish-water coastal sea. Tartu, 2010, 123 p.
175. **Eero Talts**. Photosynthetic cyclic electron transport – measurement and variably proton-coupled mechanism. Tartu, 2010, 121 p.
176. **Mari Nelis**. Genetic structure of the Estonian population and genetic distance from other populations of European descent. Tartu, 2010, 97 p.
177. **Kaarel Krjutškov**. Arrayed Primer Extension-2 as a multiplex PCR-based method for nucleic acid variation analysis: method and applications. Tartu, 2010, 129 p.
178. **Egle Köster**. Morphological and genetical variation within species complexes: *Anthyllis vulneraria* s. l. and *Alchemilla vulgaris* (coll.). Tartu, 2010, 101 p.
179. **Erki Õunap**. Systematic studies on the subfamily Sterrhinae (Lepidoptera: Geometridae). Tartu, 2010, 111 p.
180. **Merike Jõesaar**. Diversity of key catabolic genes at degradation of phenol and *p*-cresol in pseudomonads. Tartu, 2010, 125 p.
181. **Kristjan Herkül**. Effects of physical disturbance and habitat-modifying species on sediment properties and benthic communities in the northern Baltic Sea. Tartu, 2010, 123 p.
182. **Arto Pulk**. Studies on bacterial ribosomes by chemical modification approaches. Tartu, 2010, 161 p.
183. **Maria Põllupüü**. Ecological relations of cladocerans in a brackish-water ecosystem. Tartu, 2010, 126 p.
184. **Toomas Silla**. Study of the segregation mechanism of the Bovine Papillomavirus Type 1. Tartu, 2010, 188 p.
185. **Gyaneshwer Chaubey**. The demographic history of India: A perspective based on genetic evidence. Tartu, 2010, 184 p.
186. **Katrin Kepp**. Genes involved in cardiovascular traits: detection of genetic variation in Estonian and Czech populations. Tartu, 2010, 164 p.

187. **Virve Sõber**. The role of biotic interactions in plant reproductive performance. Tartu, 2010, 92 p.
188. **Kersti Kangro**. The response of phytoplankton community to the changes in nutrient loading. Tartu, 2010, 144 p.
189. **Joachim M. Gerhold**. Replication and Recombination of mitochondrial DNA in Yeast. Tartu, 2010, 120 p.
190. **Helen Tammert**. Ecological role of physiological and phylogenetic diversity in aquatic bacterial communities. Tartu, 2010, 140 p.
191. **Elle Rajandu**. Factors determining plant and lichen species diversity and composition in Estonian *Calamagrostis* and *Hepatica* site type forests. Tartu, 2010, 123 p.
192. **Paula Ann Kivistik**. ColR-ColS signalling system and transposition of Tn*4652* in the adaptation of *Pseudomonas putida.* Tartu, 2010, 118 p.
193. **Siim Sõber**. Blood pressure genetics: from candidate genes to genome-wide association studies. Tartu, 2011, 120 p.
194. **Kalle Kipper**. Studies on the role of helix 69 of 23S rRNA in the factor-dependent stages of translation initiation, elongation, and termination. Tartu, 2011, 178 p.
195. **Triinu Siibak**. Effect of antibiotics on ribosome assembly is indirect. Tartu, 2011, 134 p.
196. **Tambet Tõnissoo**. Identification and molecular analysis of the role of guanine nucleotide exchange factor RIC-8 in mouse development and neural function. Tartu, 2011, 110 p.
197. **Helin Räägel**. Multiple faces of cell-penetrating peptides – their intracellular trafficking, stability and endosomal escape during protein transduction. Tartu, 2011, 161 p.
198. **Andres Jaanus**. Phytoplankton in Estonian coastal waters – variability, trends and response to environmental pressures. Tartu, 2011, 157 p.
199. **Tiit Nikopensius**. Genetic predisposition to nonsyndromic orofacial clefts. Tartu, 2011, 152 p.
200. **Signe Värv**. Studies on the mechanisms of RNA polymerase II-dependent transcription elongation. Tartu, 2011, 108 p.
201. **Kristjan Välk**. Gene expression profiling and genome-wide association studies of non-small cell lung cancer. Tartu, 2011, 98 p.
202. **Arno Põllumäe**. Spatio-temporal patterns of native and invasive zooplankton species under changing climate and eutrophication conditions. Tartu, 2011, 153 p.
203. **Egle Tammeleht**. Brown bear (*Ursus arctos*) population structure, demographic processes and variations in diet in northern Eurasia. Tartu, 2011, 143 p.
205. **Teele Jairus**. Species composition and host preference among ectomycorrhizal fungi in Australian and African ecosystems. Tartu, 2011, 106 p.
206. **Kessy Abarenkov**. PlutoF – cloud database and computing services supporting biological research. Tartu, 2011, 125 p.

207. **Marina Grigorova**. Fine-scale genetic variation of follicle-stimulating hormone beta-subunit coding gene (*FSHB*) and its association with reproductive health. Tartu, 2011, 184 p.
208. **Anu Tiitsaar**. The effects of predation risk and habitat history on butterfly communities. Tartu, 2011, 97 p.
209. **Elin Sild**. Oxidative defences in immunoecological context: validation and application of assays for nitric oxide production and oxidative burst in a wild passerine. Tartu, 2011, 105 p.
210. **Irja Saar**. The taxonomy and phylogeny of the genera *Cystoderma* and *Cystodermella* (Agaricales, Fungi). Tartu, 2012, 167 p.
211. **Pauli Saag**. Natural variation in plumage bacterial assemblages in two wild breeding passerines. Tartu, 2012, 113 p.
212. **Aleksei Lulla**. Alphaviral nonstructural protease and its polyprotein substrate: arrangements for the perfect marriage. Tartu, 2012, 143 p.
213. **Mari Järve**. Different genetic perspectives on human history in Europe and the Caucasus: the stories told by uniparental and autosomal markers. Tartu, 2012, 119 p.
214. **Ott Scheler**. The application of tmRNA as a marker molecule in bacterial diagnostics using microarray and biosensor technology. Tartu, 2012, 93 p.
215. **Anna Balikova**. Studies on the functions of tumor-associated mucin-like leukosialin (CD43) in human cancer cells. Tartu, 2012, 129 p.
216. **Triinu Kõressaar**. Improvement of PCR primer design for detection of prokaryotic species. Tartu, 2012, 83 p.
217. **Tuul Sepp**. Hematological health state indices of greenfinches: sources of individual variation and responses to immune system manipulation. Tartu, 2012, 117 p.
218. **Rya Ero**. Modifier view of the bacterial ribosome. Tartu, 2012, 146 p.
219. **Mohammad Bahram**. Biogeography of ectomycorrhizal fungi across different spatial scales. Tartu, 2012, 165 p.
220. **Annely Lorents**. Overcoming the plasma membrane barrier: uptake of amphipathic cell-penetrating peptides induces influx of calcium ions and downstream responses. Tartu, 2012, 113 p.
221. **Katrin Männik**. Exploring the genomics of cognitive impairment: whole-genome SNP genotyping experience in Estonian patients and general population. Tartu, 2012, 171 p.
222. **Marko Prous**. Taxonomy and phylogeny of the sawfly genus *Empria* (Hymenoptera, Tenthredinidae). Tartu, 2012, 192 p.
223. **Triinu Visnapuu**. Levansucrases encoded in the genome of *Pseudomonas syringae* pv. tomato DC3000: heterologous expression, biochemical characterization, mutational analysis and spectrum of polymerization products. Tartu, 2012, 160 p.
224. **Nele Tamberg**. Studies on Semliki Forest virus replication and pathogenesis. Tartu, 2012, 109 p.

225. **Tõnu Esko**. Novel applications of SNP array data in the analysis of the genetic structure of Europeans and in genetic association studies. Tartu, 2012, 149 p.
226. **Timo Arula**. Ecology of early life-history stages of herring *Clupea harengus membras* in the northeastern Baltic Sea. Tartu, 2012, 143 p.
227. **Inga Hiiesalu**. Belowground plant diversity and coexistence patterns in grassland ecosystems. Tartu, 2012, 130 p.
228. **Kadri Koorem**. The influence of abiotic and biotic factors on small-scale plant community patterns and regeneration in boreonemoral forest. Tartu, 2012, 114 p.
229. **Liis Andresen**. Regulation of virulence in plant-pathogenic pectobacteria. Tartu, 2012, 122 p.
230. **Kaupo Kohv**. The direct and indirect effects of management on boreal forest structure and field layer vegetation. Tartu, 2012, 124 p.
231. **Mart Jüssi**. Living on an edge: landlocked seals in changing climate. Tartu, 2012, 114 p.
232. **Riina Klais**. Phytoplankton trends in the Baltic Sea. Tartu, 2012, 136 p.
233. **Rauno Veeroja**. Effects of winter weather, population density and timing of reproduction on life-history traits and population dynamics of moose (*Alces alces*) in Estonia. Tartu, 2012, 92 p.
234. **Marju Keis**. Brown bear (*Ursus arctos*) phylogeography in northern Eurasia. Tartu, 2013, 142 p.
235. **Sergei Põlme**. Biogeography and ecology of *alnus*- associated ectomycorrhizal fungi – from regional to global scale. Tartu, 2013, 90 p.
236. **Liis Uusküla**. Placental gene expression in normal and complicated pregnancy. Tartu, 2013, 173 p.
237. **Marko Lõoke**. Studies on DNA replication initiation in *Saccharomyces cerevisiae*. Tartu, 2013, 112 p.
238. **Anne Aan**. Light- and nitrogen-use and biomass allocation along productivity gradients in multilayer plant communities. Tartu, 2013, 127 p.
239. **Heidi Tamm**. Comprehending phylogenetic diversity – case studies in three groups of ascomycetes. Tartu, 2013, 136 p.
240. **Liina Kangur**. High-Pressure Spectroscopy Study of Chromophore-Binding Hydrogen Bonds in Light-Harvesting Complexes of Photosynthetic Bacteria. Tartu, 2013, 150 p.
241. **Margus Leppik**. Substrate specificity of the multisite specific pseudouridine synthase RluD. Tartu, 2013, 111 p.
242. **Lauris Kaplinski**. The application of oligonucleotide hybridization model for PCR and microarray optimization. Tartu, 2013, 103 p.
243. **Merli Pärnoja**. Patterns of macrophyte distribution and productivity in coastal ecosystems: effect of abiotic and biotic forcing. Tartu, 2013, 155 p.
244. **Tõnu Margus**. Distribution and phylogeny of the bacterial translational GTPases and the Mqsr/YgiT regulatory system. Tartu, 2013, 126 p.
245. **Pille Mänd**. Light use capacity and carbon and nitrogen budget of plants: remote assessment and physiological determinants. Tartu, 2013, 128 p.

246. **Mario Plaas**. Animal model of Wolfram Syndrome in mice: behavioural, biochemical and psychopharmacological characterization. Tartu, 2013, 144 p.
247. **Georgi Hudjašov**. Maps of mitochondrial DNA, Y-chromosome and tyrosinase variation in Eurasian and Oceanian populations. Tartu, 2013, 115 p.
248. **Mari Lepik**. Plasticity to light in herbaceous plants and its importance for community structure and diversity. Tartu, 2013, 102 p.
249. **Ede Leppik**. Diversity of lichens in semi-natural habitats of Estonia. Tartu, 2013, 151 p.
250. **Ülle Saks**. Arbuscular mycorrhizal fungal diversity patterns in boreo-nemoral forest ecosystems. Tartu, 2013, 151 p.
251. **Eneli Oitmaa**. Development of arrayed primer extension microarray assays for molecular diagnostic applications. Tartu, 2013, 147 p.
252. **Jekaterina Jutkina**. The horizontal gene pool for aromatics degradation: bacterial catabolic plasmids of the Baltic Sea aquatic system. Tartu, 2013, 121 p.
253. **Helen Vellau**. Reaction norms for size and age at maturity in insects: rules and exceptions. Tartu, 2014, 132 p.
254. **Randel Kreitsberg**. Using biomarkers in assessment of environmental contamination in fish – new perspectives. Tartu, 2014, 107 p.
255. **Krista Takkis**. Changes in plant species richness and population performance in response to habitat loss and fragmentation.Tartu, 2014, 141 p.
256. **Liina Nagirnaja**. Global and fine-scale genetic determinants of recurrent pregnancy loss. Tartu, 2014, 211 p.
257. **Triin Triisberg**. Factors influencing the re-vegetation of abandoned extracted peatlands in Estonia. Tartu, 2014, 133 p.
258. **Villu Soon**. A phylogenetic revision of the *Chrysis ignita* species group (Hymenoptera: Chrysididae) with emphasis on the northern European fauna. Tartu, 2014, 211 p.
259. **Andrei Nikonov**. RNA-Dependent RNA Polymerase Activity as a Basis for the Detection of Positive-Strand RNA Viruses by Vertebrate Host Cells. Tartu, 2014, 207 p.
260. **Eele Õunapuu-Pikas**. Spatio-temporal variability of leaf hydraulic conductance in woody plants: ecophysiological consequences. Tartu, 2014, 135 p.
261. **Marju Männiste**. Physiological ecology of greenfinches: information content of feathers in relation to immune function and behavior. Tartu, 2014, 121 p.
262. **Katre Kets**. Effects of elevated concentrations of CO_2 and O_3 on leaf photosynthetic parameters in *Populus tremuloides*: diurnal, seasonal and interannual patterns. Tartu, 2014, 115 p.
263. **Külli Lokko**. Seasonal and spatial variability of zoopsammon communities in relation to environmental parameters. Tartu, 2014, 129 p.
264. **Olga Žilina**. Chromosomal microarray analysis as diagnostic tool: Estonian experience. Tartu, 2014, 152 p.

265. **Kertu Lõhmus**. Colonisation ecology of forest-dwelling vascular plants and the conservation value of rural manor parks. Tartu, 2014, 111 p.
266. **Anu Aun**. Mitochondria as integral modulators of cellular signaling. Tartu, 2014, 167 p.
267. **Chandana Basu Mallick**. Genetics of adaptive traits and gender-specific demographic processes in South Asian populations. Tartu, 2014, 160 p.
268. **Riin Tamme**. The relationship between small-scale environmental heterogeneity and plant species diversity. Tartu, 2014, 130 p.
269. **Liina Remm**. Impacts of forest drainage on biodiversity and habitat quality: implications for sustainable management and conservation. Tartu, 2015, 126 p.
270. **Tiina Talve**. Genetic diversity and taxonomy within the genus *Rhinanthus.* Tartu, 2015, 106 p.
271. **Mehis Rohtla**. Otolith sclerochronological studies on migrations, spawning habitat preferences and age of freshwater fishes inhabiting the Baltic Sea. Tartu, 2015, 137 p.
272. **Alexey Reshchikov**. The world fauna of the genus *Lathrolestes* (Hymenoptera, Ichneumonidae). Tartu, 2015, 247 p.
273. **Martin Pook**. Studies on artificial and extracellular matrix protein-rich surfaces as regulators of cell growth and differentiation. Tartu, 2015, 142 p.
274. **Mai Kukumägi**. Factors affecting soil respiration and its components in silver birch and Norway spruce stands. Tartu, 2015, 155 p.
275. **Helen Karu**. Development of ecosystems under human activity in the North-East Estonian industrial region: forests on post-mining sites and bogs. Tartu, 2015, 152 p.
276. **Hedi Peterson**. Exploiting high-throughput data for establishing relationships between genes. Tartu, 2015, 186 p.
277. **Priit Adler**. Analysis and visualisation of large scale microarray data, Tartu, 2015, 126 p.
278. **Aigar Niglas**. Effects of environmental factors on gas exchange in deciduous trees: focus on photosynthetic water-use efficiency. Tartu, 2015, 152 p.
279. **Silja Laht**. Classification and identification of conopeptides using profile hidden Markov models and position-specific scoring matrices. Tartu, 2015, 100 p.
280. **Martin Kesler**. Biological characteristics and restoration of Atlantic salmon *Salmo salar* populations in the Rivers of Northern Estonia. Tartu, 2015, 97 p.
281. **Pratyush Kumar Das**. Biochemical perspective on alphaviral nonstructural protein 2: a tale from multiple domains to enzymatic profiling. Tartu, 2015, 205 p
282. **Priit Palta**. Computational methods for DNA copy number detection. Tartu, 2015, 130 p.
283. **Julia Sidorenko**. Combating DNA damage and maintenance of genome integrity in pseudomonads. Tartu, 2015, 174 p.
284. **Anastasiia Kovtun-Kante**. Charophytes of Estonian inland and coastal waters: distribution and environmental preferences. Tartu, 2015, 97 p.

285. **Ly Lindman**. The ecology of protected butterfly species in Estonia. Tartu, 2015, 171 p.
286. **Jaanis Lodjak**. Association of Insulin-like Growth Factor I and Corticosterone with Nestling Growth and Fledging Success in Wild Passerines. Tartu, 2016, 113 p.
287. **Ann Kraut**. Conservation of Wood-Inhabiting Biodiversity – Semi-Natural Forests as an Opportunity. Tartu, 2016, 141 p.
288. **Tiit Örd**. Functions and regulation of the mammalian pseudokinase TRIB3. Tartu, 2016, 182. p.
289. **Kairi Käiro**. Biological Quality According to Macroinvertebrates in Streams of Estonia (Baltic Ecoregion of Europe): Effects of Human-induced Hydromorphological Changes. Tartu, 2016, 126 p.
290. **Leidi Laurimaa**. *Echinococcus multilocularis* and other zoonotic parasites in Estonian canids. Tartu, 2016, 144 p.
291. **Helerin Margus**. Characterization of cell-penetrating peptide/nucleic acid nanocomplexes and their cell-entry mechanisms. Tartu, 2016, 173 p.
292. **Kadri Runnel**. Fungal targets and tools for forest conservation. Tartu, 2016, 157 p.
293. **Urmo Võsa**. MicroRNAs in disease and health: aberrant regulation in lung cancer and association with genomic variation. Tartu, 2016, 163 p.
294. **Kristina Mäemets-Allas**. Studies on cell growth promoting AKT signaling pathway – a promising anti-cancer drug target. Tartu, 2016, 146 p.
295. **Janeli Viil**. Studies on cellular and molecular mechanisms that drive normal and regenerative processes in the liver and pathological processes in Dupuytren's contracture. Tartu, 2016, 175 p.
296. **Ene Kook**. Genetic diversity and evolution of *Pulmonaria angustifolia* L. and *Myosotis laxa sensu lato* (Boraginaceae). Tartu, 2016, 106 p.
297. **Kadri Peil**. RNA polymerase II-dependent transcription elongation in *Saccharomyces cerevisiae*. Tartu, 2016, 113 p.
298. **Katrin Ruisu**. The role of RIC8A in mouse development and its function in cell-matrix adhesion and actin cytoskeletal organisation. Tartu, 2016, 129 p.
299. **Janely Pae**. Translocation of cell-penetrating peptides across biological membranes and interactions with plasma membrane constituents. Tartu, 2016, 126 p.
300. **Argo Ronk**. Plant diversity patterns across Europe: observed and dark diversity. Tartu, 2016, 153 p.
301. **Kristiina Mark**. Diversification and species delimitation of lichenized fungi in selected groups of the family Parmeliaceae (Ascomycota). Tartu, 2016, 181 p.
302. **Jaak-Albert Metsoja**. Vegetation dynamics in floodplain meadows: influence of mowing and sediment application. Tartu, 2016, 140 p.
303. **Hedvig Tamman**. The GraTA toxin-antitoxin system of *Pseudomonas putida*: regulation and role in stress tolerance. Tartu, 2016, 154 p.

304. **Kadri Pärtel**. Application of ultrastructural and molecular data in the taxonomy of helotialean fungi. Tartu, 2016, 183 p.
305. **Maris Hindrikson**. Grey wolf (*Canis lupus*) populations in Estonia and Europe: genetic diversity, population structure and -processes, and hybridization between wolves and dogs. Tartu, 2016, 121 p.
306. **Polina Degtjarenko**. Impacts of alkaline dust pollution on biodiversity of plants and lichens: from communities to genetic diversity. Tartu, 2016, 126 p.
307. **Liina Pajusalu**. The effect of CO_2 enrichment on net photosynthesis of macrophytes in a brackish water environment. Tartu, 2016, 126 p.
308. **Stoyan Tankov**. Random walks in the stringent response. Tartu, 2016, 94 p.
309. **Liis Leitsalu**. Communicating genomic research results to population-based biobank participants. Tartu, 2016, 158 p.
310. **Richard Meitern**. Redox physiology of wild birds: validation and application of techniques for detecting oxidative stress. Tartu, 2016, 134 p.
311. **Kaie Lokk**. Comparative genome-wide DNA methylation studies of healthy human tissues and non-small cell lung cancer tissue. Tartu, 2016, 127 p.
312. **Mihhail Kurašin**. Processivity of cellulases and chitinases. Tartu, 2017, 132 p.
313. **Carmen Tali**. Scavenger receptors as a target for nucleic acid delivery with peptide vectors. Tartu, 2017, 155 p.
314. **Katarina Oganjan**. Distribution, feeding and habitat of benthic suspension feeders in a shallow coastal sea. Tartu, 2017, 132 p.
315. **Taavi Paal**. Immigration limitation of forest plants into wooded landscape corridors. Tartu, 2017, 145 p.
316. **Kadri Õunap**. The Williams-Beuren syndrome chromosome region protein WBSCR22 is a ribosome biogenesis factor. Tartu, 2017, 135 p.
317. **Riin Tamm**. In-depth analysis of factors affecting variability in thiopurine methyltransferase activity. Tartu, 2017, 170 p.
318. **Keiu Kask**. The role of RIC8A in the development and regulation of mouse nervous system. Tartu, 2017, 184 p.
319. **Tiia Möller**. Mapping and modelling of the spatial distribution of benthic macrovegetation in the NE Baltic Sea with a special focus on the eelgrass *Zostera marina* Linnaeus, 1753. Tartu, 2017, 162 p.
320. **Silva Kasela**. Genetic regulation of gene expression: detection of tissue- and cell type-specific effects. Tartu, 2017, 150 p.
321. **Karmen Süld**. Food habits, parasites and space use of the raccoon dog *Nyctereutes procyonoides*: the role of an alien species as a predator and vector of zoonotic diseases in Estonia. Tartu, 2017, p.
322. **Ragne Oja**. Consequences of supplementary feeding of wild boar – concern for ground-nesting birds and endoparasite infection. Tartu, 2017, 141 p.
323. **Riin Kont**. The acquisition of cellulose chain by a processive cellobiohydrolase. Tartu, 2017, 117 p.
324. **Liis Kasari**. Plant diversity of semi-natural grasslands: drivers, current status and conservation challenges. Tartu, 2017, 141 p.

325. **Sirgi Saar**. Belowground interactions: the roles of plant genetic relatedness, root exudation and soil legacies. Tartu, 2017, 113 p.
326. **Sten Anslan**. Molecular identification of Collembola and their fungal associates. Tartu, 2017, 125 p.
327. **Imre Taal**. Causes of variation in littoral fish communities of the Eastern Baltic Sea: from community structure to individual life histories. Tartu, 2017, 118 p.
328. **Jürgen Jalak**. Dissecting the Mechanism of Enzymatic Degradation of Cellulose Using Low Molecular Weight Model Substrates. Tartu, 2017, 137 p.
329. **Kairi Kiik**. Reproduction and behaviour of the endangered European mink (*Mustela lutreola*) in captivity. Tartu, 2018, 112 p.
330. **Ivan Kuprijanov**. Habitat use and trophic interactions of native and invasive predatory macroinvertebrates in the northern Baltic Sea. Tartu, 2018, 117 p.
331. **Hendrik Meister**. Evolutionary ecology of insect growth: from geographic patterns to biochemical trade-offs. Tartu, 2018, 147 p.
332. **Ilja Gaidutšik**. Irc3 is a mitochondrial branch migration enzyme in *Saccharomyces cerevisiae*. Tartu, 2018, 161 p.
333. **Lena Neuenkamp**. The dynamics of plant and arbuscular mycorrhizal fungal communities in grasslands under changing land use. Tartu, 2018, 241 p.
334. **Laura Kasak**. Genome structural variation modulating the placenta and pregnancy maintenance. Tartu, 2018, 181 p.
335. **Kersti Riibak**. Importance of dispersal limitation in determining dark diversity of plants across spatial scales. Tartu, 2018, 133 p.
336. **Liina Saar**. Dynamics of grassland plant diversity in changing landscapes. Tartu, 2018, 206 p.
337. **Hanna Ainelo**. Fis regulates *Pseudomonas putida* biofilm formation by controlling the expression of *lapA*. Tartu, 2018, 143 p.
338. **Natalia Pervjakova**. Genomic imprinting in complex traits. Tartu, 2018, 176 p.
339. **Andrio Lahesaare**. The role of global regulator Fis in regulating the expression of *lapF* and the hydrophobicity of soil bacterium *Pseudomonas putida*. Tartu, 2018, 124 p.
340. **Märt Roosaare**. *K*-mer based methods for the identification of bacteria and plasmids. Tartu, 2018, 117 p.
341. **Maria Abakumova**. The relationship between competitive behaviour and the frequency and identity of neighbours in temperate grassland plants. Tartu, 2018, 104 p.
342. **Margus Vilbas**. Biotic interactions affecting habitat use of myrmecophilous butterflies in Northern Europe. Tartu, 2018, 142 p.
343. **Liina Kinkar**. Global patterns of genetic diversity and phylogeography of *Echinococcus granulosus* sensu stricto – a tapeworm species of significant public health concern. Tartu, 2018, 147 p.

344. **Teivi Laurimäe**. Taxonomy and genetic diversity of zoonotic tapeworms in the species complex of *Echinococcus granulosus* sensu lato. Tartu, 2018, 143 p.
345. **Tatjana Jatsenko**. Role of translesion DNA polymerases in mutagenesis and DNA damage tolerance in Pseudomonads. Tartu, 2018, 216 p.
346. **Katrin Viigand**. Utilization of α-glucosidic sugars by *Ogataea* (*Hansenula*) *polymorpha*. Tartu, 2018, 148 p.
347. **Andres Ainelo**. Physiological effects of the *Pseudomonas putida* toxin grat. Tartu, 2018, 146 p.
348. **Killu Timm**. Effects of two genes (DRD4 and SERT) on great tit (*Parus major*) behaviour and reproductive traits. Tartu, 2018, 117 p.
349. **Petr Kohout**. Ecology of ericoid mycorrhizal fungi. Tartu, 2018, 184 p.
350. **Gristin Rohula-Okunev**. Effects of endogenous and environmental factors on night-time water flux in deciduous woody tree species. Tartu, 2018, 184 p.
351. **Jane Oja**. Temporal and spatial patterns of orchid mycorrhizal fungi in forest and grassland ecosystems. Tartu, 2018, 102 p.
352. **Janek Urvik**. Multidimensionality of aging in a long-lived seabird. Tartu, 2018, 135 p.
353. **Lisanna Schmidt**. Phenotypic and genetic differentiation in the hybridizing species pair *Carex flava* and *C. viridula* in geographically different regions. Tartu, 2018, 133 p.
354. **Monika Karmin**. Perspectives from human Y chromosome – phylogeny, population dynamics and founder events. Tartu, 2018, 168 p.
355. **Maris Alver**. Value of genomics for atherosclerotic cardiovascular disease risk prediction. Tartu, 2019, 148 p.
356. **Lehti Saag**. The prehistory of Estonia from a genetic perspective: new insights from ancient DNA. Tartu, 2019, 171 p.
357. **Mari-Liis Viljur**. Local and landscape effects on butterfly assemblages in managed forests. Tartu, 2019, 115 p.
358. **Ivan Kisly**. The pleiotropic functions of ribosomal proteins eL19 and eL24 in the budding yeast ribosome. Tartu, 2019, 170 p.
359. **Mikk Puustusmaa**. On the origin of papillomavirus proteins. Tartu, 2019, 152 p.
360. **Anneliis Peterson**. Benthic biodiversity in the north-eastern Baltic Sea: mapping methods, spatial patterns, and relations to environmental gradients. Tartu, 2019, 159 p.
361. **Erwan Pennarun**. Meandering along the mtDNA phylogeny; causerie and digression about what it can tell us about human migrations. Tartu, 2019, 162 p.
362. **Karin Ernits**. Levansucrase Lsc3 and endo-levanase BT1760: characterization and application for the synthesis of novel prebiotics. Tartu, 2019, 217 p.
363. **Sille Holm**. Comparative ecology of geometrid moths: in search of contrasts between a temperate and a tropical forest. Tartu, 2019, 135 p.

364. **Anne-Mai Ilumäe**. Genetic history of the Uralic-speaking peoples as seen through the paternal haplogroup N and autosomal variation of northern Eurasians. Tartu, 2019, 172 p.
365. **Anu Lepik**. Plant competitive behaviour: relationships with functional traits and soil processes. Tartu, 2019, 152 p.
366. **Kunter Tätte**. Towards an integrated view of escape decisions in birds under variable levels of predation risk. Tartu, 2020, 172 p.
367. **Kaarin Parts**. The impact of climate change on fine roots and root-associated microbial communities in birch and spruce forests. Tartu, 2020, 143 p.
368. **Viktorija Kukuškina**. Understanding the mechanisms of endometrial receptivity through integration of 'omics' data layers. Tartu, 2020, 169 p.
369. **Martti Vasar**. Developing a bioinformatics pipeline gDAT to analyse arbuscular mycorrhizal fungal communities using sequence data from different marker regions. Tartu, 2020, 193 p.
370. **Ott Kangur**. Nocturnal water relations and predawn water potential disequilibrium in temperate deciduous tree species. Tartu, 2020, 126 p.