DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
354

MONIKA KARMIN

Perspectives from human Y chromosome – phylogeny, population dynamics and founder events

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
354

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
354

MONIKA KARMIN

Perspectives from human Y chromosome – phylogeny, population dynamics and founder events

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 Molecular Biology on 19th of October, 2018 by the Council of the Institute of Molecular and Cell Biology, Faculty of Science and Technology, University of Tartu.

Supervisors: Mait Metspalu, PhD, Senior Research Fellow of Modern Population Genetics, Institute of Genomics, University of Tartu, Estonia

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

Prof Toomas Kivisild, PhD, Department of Human Genetics, KU Leuven, Belgium
Senior Research Fellow of Population Genetics, Institute of Genomics, University of Tartu, Estonia

Ene Metspalu, PhD, Senior Research Fellow of Population Genetics, Institute of Genomics, University of Tartu, Estonia

Opponent: Prof Agnar Helgason, PhD, Department of Anthropology, University of Iceland, Senior Research Scientist in Biological Anthropology, deCODE Genetics, Iceland

Commencement: Room No 105, 23B Riia St, Tartu, on December 10th, 2018, at 10:15.

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

ISSN 1024-6479
ISBN 978-9949-77-905-5 (print)
ISBN 978-9949-77-906-2 (pdf)

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Tables:

Figures:

LIST OF ORIGINAL PUBLICATIONS

I

Karmin M, Saag L, Vicente M, Wilson Sayres MA, Järve M, Gerst Talas U, Rootsi S, Ilumäe A-M, Mägi R, Mitt M, Pagani L, Puurand T, Faltyskova Z, Clemente F, Cardona A, Metspalu E, Sahakyan H, Yunusbayev B, Hudjashov G, DeGiorgio M, Loogväli E-L, Eichstaedt C, Eelmets M, Chaubey G, Tambets K, Litvinov S, Mormina M, Xue Y, Ayub Q, Zoraqi G, Korneliussen TS, Akhatova F, Lachance J, Tishkoff S, Momynaliev K, Ricaut F, Kusuma P, Razafindrazaka H, Pierron D, Cox MP, Sultana GNN, Willerslev R, Muller C, Westaway M, Lambert D, Skaro V, Kovačević L, Turdikulova S, Dalimova D, Khusainova R, Trofimova N, Akhmetova V, Khidiyatova I, Lichman DV, Isakova J, Pocheshkhova E, Sabitov Z, Barashkov NA, Nymadawa P, Mihailov E, Seng JWT, Evseeva I, Migliano AB, Abdullah S, Andriadze G, Primorac D, Atramentova L, Utevska O, Yepiskoposyan L, Marjanović D, Kushniarevich A, Behar DM, Gilissen C, Vissers L, Veltman JA, Balanovska E, Derenko M, Malyarchuk B, Metspalu A, Fedorova S, Eriksson A, Manica A, Mendez FL, Karafet TM, Veeramah KR, Bradman N, Hammer MF, Osipova LP, Balanovsky O, Khusnutdinova EK, Johnsen K, Remm M, Thomas MG, Tyler-Smith C, Underhill PA, Willerslev E, Nielsen R, Metspalu M, Villems R, Kivisild T. 2015**. A recent bottleneck of Y chromosome diversity coincides with a global change in culture.** *Genome Research* 25(4):459–466.

II

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW Jr, Orlando L, Metspalu E, **Karmin M**, Tambets K, Rootsi S, Mägi R, Campos PF, Balanovska E, Balanovsky O, Khusnutdinova E, Litvinov S, Osipova LP, Fedorova SA, Voevoda MI, DeGiorgio M, Sicheritz-Ponten T, Brunak S, Demeshchenko S, Kivisild T, Villems R, Nielsen R, Jakobsson M, Willerslev E. 2014. **Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans.** *Nature*, 505(7481), 87–91.

III

Rai N, Chaubey G, Tamang R, Pathak AK, Singh VK, **Karmin M**, Singh M, Rani DS, Anugula S, Yadav BK, Singh A, Srinivasagan R, Yadav A, Kahsyap M, Narvariya S, Reddy AG, VanDriem G, Underhill PA, Villems R, Kivisild T, Singh L, Thangaraj K. 2012. **The Phylogeography of Y-Chromosome Haplogroup H1a1a-M82 Reveals the Likely Indian Origin of the European Romani Populations.** *PLoS ONE*, 7, e48477.

IV

Behar DM, Saag L, **Karmin M**, Gover MG, Wexler JD, Fernanda L, Greenspan E, Kushniarevich A, Davydenko O, Sahakyan H, Yepiskoposyan L, Boattini A, Sarno S, Pagani L, Carmi S, Tzur S, Metspalu E, Bormans C, Skorecki K,

Metspalu M, Rootsi S, Villems R. 2017. **The genetic variation in the R1a clade among the Ashkenazi Levites' Y chromosome.** *Scientific Reports*, Nov 2;7(1):14969.

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

REF I: I developed and implemented filtering steps and analysed the data, interpreted results and co-wrote the manuscript.

REF II: I performed the phylogenetic analyses for chrY data, interpreted the results and provided input for the writing of the paper.

REF III: I interpreted results and was involved in the writing of the paper.

REF IV: I performed the phylogenetic analyses and age calculations, interpreted results and co-wrote the paper.

ABBREVIATIONS

AMH	anatomically modern human
AMP	ampliconic region of Y chromosome
BAC	bacterial artificial chromosomes
B/CE	before common era
BSP	Bayesian skyline plot
chrX	X chromosome
chrY	Y chromosome
F_{ST}	fixation index, a special case of Wright's *F*-statistics that measures population differentiation due to genetic structure
IBD	identity by descent
ISOGG	International Society of Genetic Genealogy, https://isogg.org/
hg	haplogroup
K/MYA	thousand (kilo)/ million years ago
LGM	last glacial maximum
lncRNA	long non-coding RNA
MSY/NRY	male specific region of Y chromosome, non-recombining part of Y chromosome
mtDNA	mitochondrial DNA
$N_e/N_f/N_m$	effective population size, female effective population size, male effective population size
NORF	not annotated open reading frame
OOA	Out-of-Africa
PAR	pseudoautosomal region of Y chromosome
SNV	single nucleotide variant (aka SNP, single nucleotide polymorphism)
SRS/NGS	short read sequencing technology, next generation sequencing technology
SRY	sex determining locus Y
STR	short tandem repeat
(T)MRCA	(time to the) most recent common ancestor
TSPY	testis-specific protein on Y chromosome
XDG	X-degenerated region
XTR	X-transposed region
YCC	Y Chromosome Consortium

1. INTRODUCTION

Human demographic history has continued to fascinate both the scientists and the general public. For many decades the consensus view on the demographic history of humans derived from archaeological and genetic evidence was that our ancestors, Anatomically Modern Humans (AMH), emerged as a species around 200 thousand years ago (KYA) from a single group in Eastern Africa (Jobling *et al*. 2014). Only very recently, new archaeological findings from Morocco have pushed the age of humankind back more than a half, to about 315 KYA (Hublin *et al*. 2017). This finding also demonstrates that early AMH inhabited a much wider area within Africa than previously thought [reviewed in (Scerri *et al*. 2018)]. The consensus was that around 50–70 KYA a small founding population of AMH moved out from Africa to colonize the rest of the world, known as the Out-of-Africa (OOA) migration or OOA bottleneck. This founding population gave rise to all the human populations and their genetic diversity outside Africa containing only a subset of genetic lineages present in Africa at that time. The subsequent migrations took people to Europe (43–46 KYA), Australia (~40 KYA), reaching Americas (~15 KYA), and Eastern Polynesia the latest (~1 KYA) (Jobling *et al*. 2014).

Recombination shuffles variation at most loci in the human genome in each generation, so that each part of an individual genome can reflect different aspects of population histories of multiple genealogical ancestors of that individual. Genome-scale analyses face, thus, a difficult challenge to disentangle the demographic history of the populations that the ancestors of the studied individuals represent. Recent massive analyses of whole genome sequencing data from modern individuals have confirmed and added much detail to the previously vested picture of human population demographic history, but also challenged the single origin, OOA migration and complete replacement theory. Genomic evidence shows that the first splits in human populations predate the consensus of ~200 KYA (Schlebusch *et al*. 2017; Skoglund *et al*. 2017) and that humans evolved in subdivided populations within Africa (Scerri *et al*. 2018). Another line of data proposes an additional earlier OOA migration that have left subtle traces only into the genomes of contemporary people from Papua New Guinea (Pagani *et al*. 2016; Rasmussen *et al*. 2011). Also, the ever more abundant genomic sequences from ancient specimen have revealed that at its early stages the human population not only co-inhabited several regions with their cousins from the hominine line, namely *Homo neanderthalensis* (Green *et al*. 2010; Prüfer *et al*. 2014) and *Homo sapiens ssp. Denisova* (Meyer *et al*. 2012; Reich *et al*. 2010; Slon *et al*. 2017), but also repeatedly admixed with them (Browning *et al*. 2018; Sankararaman *et al*. 2016; Slon *et al*. 2018; Wall and Yoshihara Caldeira Brandt 2016).

At the same time when the analyses of whole human genomes are revolutionising the view on the demographic history of humankind, much has changed for how a person can study their own genealogy. Personal molecular

genealogy is widely spread, with both the genome-wide and haploid markers, the maternally inherited mitochondrial genome (mtDNA) and the paternally inherited Y chromosome (chrY), being commonly analysed. The capacity to approach personal genealogy with molecular tools provides a wealth of data studied with great enthusiasm; analyses of these personal genomes bring together fruitful collaboration of academic and citizen scientists providing new insight on a population level as well (Balanovsky *et al*. 2017; Jobling and Tyler-Smith 2017).

This thesis touches upon aspects of human demographic history learned from the analyses of different units of human genome, with the main focus on the Y chromosome. It gives an overview of the chrY functions and evolution, distribution of extant and ancient lineages and a summary of the history of European Roma and Ashkenazi Levites, two examples of populations with past founder events. Uniparentally inherited parts of the genomes open a unique possibility to study sex-specific aspects of human demographic history. Fine-tuning the chrY data that so far suffered from much lower marker density to its currently highest possible resolution in a globally distributed sample is the starting point for the first study. This allows the construction of a dated phylogenetic tree of global paternal lineages and the comparison of the dynamics of male and female effective population size (N_e) (REF I). Then, an individual whole genome sequence, including the Y chromosome, of a boy who died ~24,000 years ago near lake Baikal adds new insight to the ancestry of contemporary Western Eurasian and American populations (REF II).

Distinct patterns of chrY diversity provide evidence of founder events. The thesis includes two case studies that focus on two such lineages prevalent among specific populations. In particular our third study examines the distribution of the paternal lineage H1a1a within India and its implications to the origin of European Roma (REF III). Finally, in the fourth study the capture and sequencing of chrY gives a detailed overview of the phylogeny, origin and spread of the paternal lineage R1a-M582 (REF IV). This haplogroup is prevalent among the Ashkenazi Levites, the Jewish priestly class and it is analysed in comparison to other frequent male lineages among the Ashkenazi Jews.

2. LITERATURE OVERVIEW

The literature overview gives brief background information about chrY studies, a concise picture of the chrY evolution, structure and function. Then, literature on more specific topics relevant to the studies included in the thesis is discussed.

2.1. Human Y chromosome – evolution, structure and function

In humans the genetic sex of an individual is determined by the twenty third pair of chromosomes. Human males have typically inherited one copy of chrY with male sex determining genes from the father and one copy of chromosome X (chrX) from the mother. These chromosomes form a homologous pair. Females typically carry two copies of homologous X chromosomes, one inherited from mother and the other from father. Phenotypic sex determination is more complex, though, than mere chromosomal sex. When there are defects in the sex determining genes on chrY or in some autosomal genes, including those coding for androgen sensitivity, then individuals with chrY can develop as females (Délot *et al*. 2017). During meiosis chrY pairs with chrX, which, contrary to other homologous chromosome pairs in human genome, is different in size and has much reduced sequence similarity with its recombination partner. As a result just small distinct parts located in the ends of these chromosomes recombine. This keeps the majority of chrY as a single, male-specific locus – a block of DNA that does not recombine with its homologue and is passed on in male lineage only (Jobling and Tyler-Smith 2003; Skaletsky *et al*. 2003).

The research of human Y chromosome has been divided into three distinct eras (Skaletsky *et al*. 2003). First, the scholars in the beginning of 20th century focused on the Mendelian diseases, reporting by 1950 at least 17 Y-linked traits (reviewed in (Stern 1957)). These findings were systematically confuted in the next stage (Stern 1957) and the case of hairy ears more recently (Lee *et al*. 2004). During the second stage chrY was thought to be genetically barren, despite the findings of the essential male sex-determining genes on that very chromosome (Ford and Jones 1959; Jacobs and Strong 1959). When in 1967 Ohno proposed that mammalian chrY and chrX evolved from an ancestral pair of autosomes, genes of which chrX had mostly retained, it engraved the general idea of chrY as a decayed version of chrX, with a very limited genetic content and even on the verge of complete degradation (Charlesworth 1991; Graves 2006; Ohno 1966). The current, third era starting in 1980s, is marked by a vast array of molecular technologies applied to studying the chrY, revealing its complex biological functions, molecular structure, evolution and distribution of its variants in the present and past human populations (Bachtrog 2013; Bellott *et al*. 2014; Burgoyne 1998; Cortez *et al*. 2014; Hughes and Rozen 2012; Jobling, Mark A.;

Tyler-Smith 2003; Kuroda-Kawaguchi *et al*. 2001; Lahn and Page 1997; Page *et al*. 1984; Sargent *et al*. 1996; Skaletsky *et al*. 2003; Sun and Heitman 2012; Tilford *et al*. 2001; Underhill *et al*. 2001; Underhill and Kivisild 2007).

The (mammalian) sex chromosome

ChrY as the male sex-determining chromosome is common to most mammals (Cortez *et al*. 2014; Kashimada and Koopman 2010). The mammalian sex chromosomes originated from a pair of autosomal chromosomes, one of which developed the sex determining locus (*SRY*), followed by several inversion events that led to the almost complete recombination arrest between the original autosomes (Figure 1) (Graves 1995; Katsura *et al*. 2012; Lahn and Page 1999; Waters *et al*. 2001). Human sex chromosomes bear two major evolutionary layers – the older parts retained from the ancestral autosomes and shared between all therian mammals (X-conserved region), and Y- and X-added regions (YAR/XAR) from an autosomal translocation 80–130 MYA common to all eutherian mammals (Graves 1995; Waters *et al*. 2001). Through the step-wise loss of recombination during millions of years of evolution chrY decayed and preserved only three percent of its original gene content. The chrY decay was initially rapid but reached a stasis over the last 25 million years (Bellott *et al*. 2014; Hallast and Jobling 2017; Hughes *et al*. 2012).

The gametologues, homologous genes on the chrX and chrY, diverged from each other in a step-wise manner resulting in at least five distinct evolutionary strata (Bellott *et al*. 2014; Hughes *et al*. 2012; Lahn and Page 1999; Skaletsky *et al*. 2003). With each stratum (Figure 1) the part of chrY that does not undergo recombination increased at the expense of the recombining parts (Hughes *et al*. 2012). The lack of recombination relaxes the demand for genetic homology between relevant regions and this caused most the genes on chrY to become non-functional through the accumulation of mutations. Genes became pseudo-genes, diverged beyond recognition or were completely lost. On the other hand, a specific set of dosage-sensitive ancestral genes were retained on chrY and the genes important for male fertility accumulated (Bellott *et al*. 2014; Cortez *et al*. 2014; Skaletsky *et al*. 2003; Wilson Sayres and Makova 2013). In females the inactivation of one homologue on chrX evolved as a response to the loss of a gene (or its function) on chrY. This compensates for the resulting dosage difference of gene expression between males and females (Bishop *et al*. 1984; Charlesworth 1978; Graves 1995; Lahn *et al*. 2001; Ohno 1966; Waters *et al*. 2001).

The non-recombining parts of chrY have gone through numerous structural changes (Archidiacono *et al*. 1998; Hallast and Jobling 2017; Hughes *et al*. 2010; Page *et al*. 1984; Ross *et al*. 2005; Skaletsky *et al*. 2003). The entire complex evolution of the sex chromosome has resulted in the sequence and structure of the contemporary human Y chromosomes.

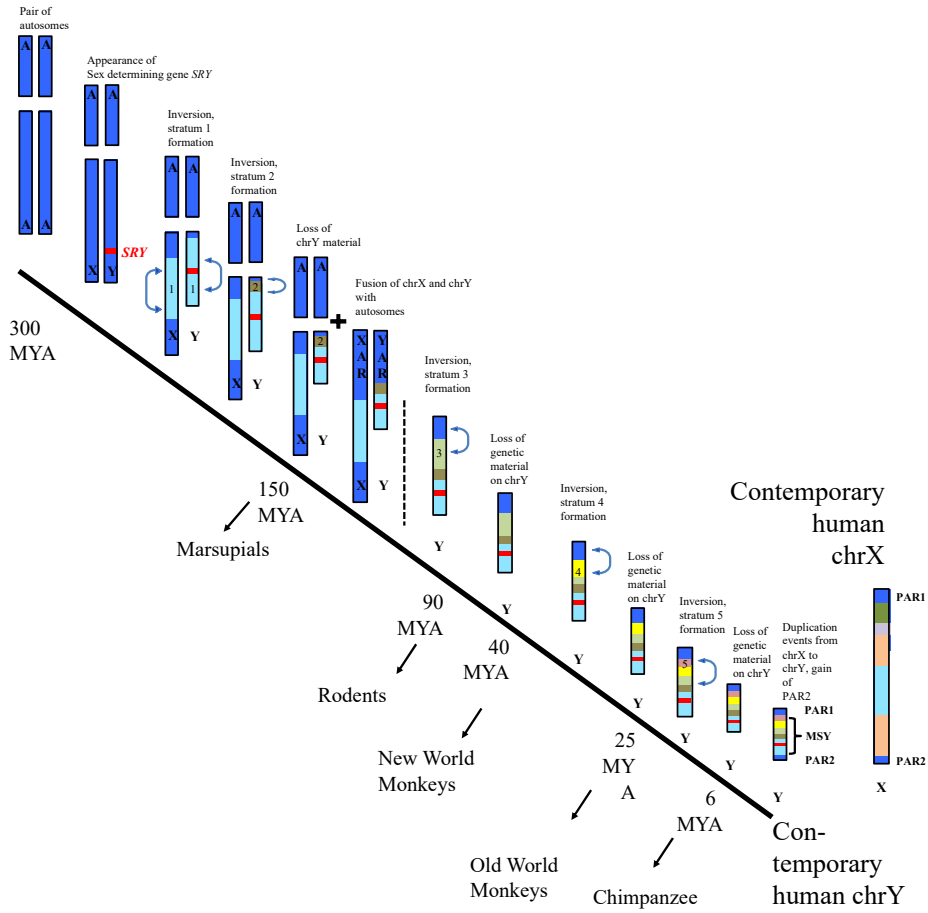

Figure 1. Schematic overview of human sex chromosome evolution from the ancestral autosomes. A sex-determining gene *SRY* evolves on one of the ancestral autosomes (A) that later becomes chrY (Y). Inversions (shown with two-sided arrows) on both of the ancestral sex chromosomes and fusion of chrX and chrY with a pair of autosomes (shown with plus sign) lead to almost complete recombination arrest between chrY and chrX. Key time points of chrY evolution are shown on the diagonal. For chrX only the initial evolutionary events until the fusion with autosome are shown (up to the dashed line on the figure). Contemporary chrX has five evolutionary strata. Inversions take place on both chromosomes, there is substantial loss of sequence in chrY but not in chrX. *Abbreviations*: *SRY* – male sex determining gene, PAR – pseudoautosomal region, MSY – male-specific region of chrY, XAR/YAR – 'X-added region'/'Y-added region' – large translocation from a pair of autosomes to the sex chromosomes in the lineage of eutherian mammals (Graves 1995). Colours denote the evolutionary strata on chrY and chrX, they are schematic and not shown to scale. During the evolution of chrY MSY expanded with each stratum at the expense of the PAR (blue) region whereas the fusion with other autosomes extended the PAR. PAR2 is present only in contemporary humans and was gained from duplication events from chrX to chrY during last few million years. Figure adapted from Figure 7 (Hughes, Skaletsky and Page, 2012), ANNUAL REVIEWS and Figure 3 (Bellott *et al*. 2014), Springer Nature.

Sequence classes of human Y chromosome

Continuing on the previous efforts Skaletsky and colleagues published a reference sequence of chrY from one man (Skaletsky *et al*. 2003). The seminal study for all further analyses characterizes extensively the structure and sequence of chrY (Kuroda-Kawaguchi *et al*. 2001; Skaletsky *et al*. 2003; Sun *et al*. 1999; Tilford *et al*. 2001). With its ~59 mega bases (Mb) chrY is one of the shortest chromosomes of the genome and the sequence is classified into *pseudoautosomal* (PAR1 and PAR2) and male specific chrY (MSY). In turn, MSY is a mosaic of distinct sequence classes – *X-transposed*, *X-degenerate*, *ampliconic*, and *heterochromatic* (Figure 2; Table 1) (Hughes and Rozen 2012; Skaletsky *et al*. 2003).

The crossover between chrX and chrY in male meiosis occurs between homologous regions in the very ends of both of the chromosomes, called pseudoautosomal regions PAR1 and PAR2 (Figure 1 and Figure 2a, Table1). PAR1 is at the tip of the short arms of chrY and chrX (Yp and Xp) spanning 2.7 Mb, gene loci in this region are present in two copies in both males and females and are not subject to dosage compensation in females. The second pseudoautosomal region, PAR2, with its 330 thousand bases (Kb) is at the tip of long arms of chrX and chrY (Yq and Xq) (Mangs and Morris 2007; Ross *et al*. 2005; Skaletsky *et al*. 2003). Human chrY is the only Y chromosome characterized thus far with two pseudoautosomal regions. PAR1 originates from the YAR translocated approximately 80–130 MYA (Waters *et al*. 2001). PAR2 originates from at least two duplication events from chrX into chrY that took place during the last few million years after the divergence of human and chimpanzee (Charchar *et al*. 2003). Some of the genes in PAR2 are subject to dosage compensation in females, presumably reflecting their origin from chrX (Freije *et al*. 1992; Ross *et al*. 2005).

Majority of chrY, 95% of its length, does not recombine with chrX, but see also (Cotter *et al*. 2016). Therefore this region was originally called the 'non-recombining region of chrY' (NRY) (Hallast and Jobling 2017; Lahn and Page 1997). After the discovery of abundant interchromosomal recombination it was renamed to 'male-specific region of chrY' or MSY (Figure 1, Figure 2 a.) (Rozen *et al*. 2003).

Table 1. Sequence classes of human Y chromosome, their length and characteristics. Colours as in Figure 2, adapted from (Skaletsky *et al.* 2003).

Eu-/hetero-chromatin	MSY/ non-MSY	Sequence class	Characteristics	Evolutionary origins	Distribution	Aggregate length (Mb)	Coding genes; expression	No. of non-coding transcription units
Euchromatin (25 Mb)	MSY	X-transposed (XTR)	99% identity to chrX	Single transposition form chrX 3–4.7 MYA	2 blocks, Yp	3.4	2	0
		X-degenerate (XDG)	Single-copy genes or pseudogenes of chrX-linked genes	Relics of ancient autosomes	8 blocks, Yp and Yq	8.6	16; most ubiquitous	4
		Ampliconic (AMP)	Highly similar to other MSY sequences	Diverse sources and sequentially amplified	7 blokcs, Yp and Yq	10.2	60 (9 families); mainly in testes	74 (9 single-copy; 65 in 15 families), expressed mainly in testes
	non-MSY	Pseudo-autosomal (PAR)	Regular crossing-over with chrX	PAR1 – early autosomal translocation 80–130 MYA; PAR2 – chrX duplications 2–3 MYA	2 blocks on the extreme ends of Yp/Xp and Yq/Xq	2.7+0.3	24+4; ubiquitous	NA
Hetero-chromatin	MSY	Hetero-chromatic (YHET)	No transcription units, six sequence classes	NA	3 blocks (centromeric, 40 Mb on Yq and 400 kb on proximal Yq)	variable	No transcribed elements	No transcribed elements

MSY contains both euchromatic and heterochromatic sequences. The heterochromatic sequence is in several discrete blocks, the largest of which is the ~40 Mb on the long arm of the chrY. In addition to this there is the centromeric block and an island of heterochromatin that interrupts the euchromatin on the proximal long arm of chrY (Figure 2). The third block spans ~400 Kb and includes 3,000 tandem repeats of 125 base pairs (bp). In total the heterochromatin of MSY encompaases at least six distinct sequence species, each of which form long, homogeneous tandem arrays (Skaletsky *et al*. 2003). The euchromatin of human MSY holds interspersed X-transposed, X-degenerate and ampliconic regions (Figure 2 b.).

X-transposed region (XTR) (Figure 2 b., Table 1) originates from a 3–4.7 MY old transposition event from chrX to chrY with 3.38 Mb of the original sequence retained. The sequence identity between two chromosomes is 98.78%, when excluding indels (Hughes *et al*. 2010; Page *et al*. 1984; Ross *et al*. 2005; Skaletsky *et al*. 2003). There are substantial changes to the XTR since the transposition event. An inversion has separated a 200-Kb section from the bulk of the XTR and the 540 Kb from the original chrX sequence have been subsequently deleted (Mumm *et al*. 1997; Ross *et al*. 2005; Sargent *et al*. 1996; Skaletsky *et al*. 2003). XTR region has the highest density of interspersed repeat elements on the MSY euchromatin, for example long interspersed nuclear element 1 (LINE1) sequences account for 36% of all XTR sequence (Skaletsky *et al*. 2003).

Most prominent euchromatic sequence classes of chrY are *X-degenerate* (XDG) and *ampliconic* (AMP) regions (Figure 2 b., Table 1), which have distinct evolutionary history and gene repertoires (Lahn and Page 1997; Skaletsky *et al*. 2003). The 8.6 Mb of X-degenerate sequence make up a decayed version of chrX with sparse single-copy homologues to chrX-linked genes or pseudogenes with nucleotide sequence identity from 60% to 96%. These genes are the relics of the ancestral autosomes from which the sex chromosomes evolved, most of them ubiquitously expressed in the body (Skaletsky *et al*. 2003).

Scattered across the euchromatic long arm and proximal short arm of chrY with combined length of 10.2 Mb are the seven *ampliconic segments* (Figure 2 b., Table 1). They constitute 45% of the euchromatic sequence of chrY and belong also to the most gene-rich third type of the sequence. These segments are made up of long MSY-specific repeat units or amplicons that contain multicopy genes which belong to nine families and show predominantly testis-specific expression. 60% of these segments have as much as 99.9 % sequence identity over tens or hundreds of kilobases of other sequences in MSY. There is an abundance of similarities within and between ampliconic segments on both arms of the Y chromosome. Sequences within ampliconic segments are arranged in palindromes, inverted repeats, and tandem arrays. There are 8 palindromes that take up a quarter (5.7 Mb) of all the euchromatic sequence of chrY (Figure 2 d.). Palindromes are long, their arm length ranging from 9 Kb to 1.45 Mb and the longest, P1, makes up to the total of 2.9 Mb. All palindromes contain also a central spacer sequence that is not repeated, but sequence similarity of arms is 99.94–99.99%.

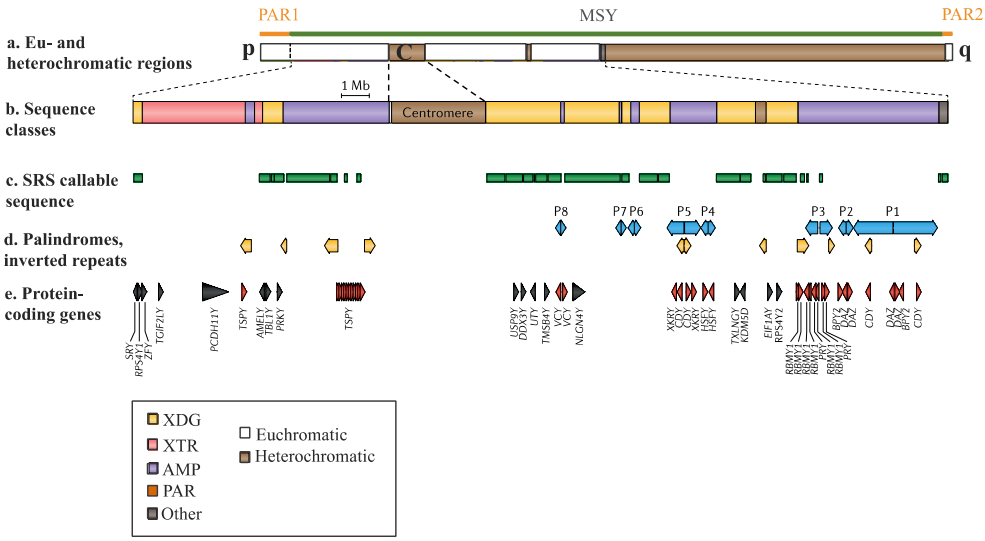

Figure 2. Schematic overview of human Y chromosomal sequence classes, short read sequencing (SRS) callable regions, repetitive sequence content and genes. a. *Eu- and heterochromatic regions.* The horizontal bar represents human chrY, p – small arm, q long arm, C – centromere. *Colours* as shown on the legend. The coloured line on top shows boundaries of regions that recombine with chrX: orange – pseudoautosomal region (PAR) where crossing over with chrX occurs, green – the male specific region of chrY (MSY) that is passed through generations without recombination with chrX. **b. *Sequence classes within the mainly euchromatic region of MSY.*** *Colours* as shown on the legend. *Abbreviations* (see text for more details): X-degenerate (XDG); X-transposed (XTR); ampliconic (AMP). Close to the centromere there are other types of sequence. **c. *SRS callable regions.*** Green bars show the regions (total ~10 Mb) with most unique and least repetitive content where mapping of short reads occurs with high accuracy resulting in high confidence base-calling (Poznik *et al*. 2013). **d. *Inverted repeats and palindromes.*** Extensive amount of sequence in chrY is highly repetitive, arrows show the direction of inverted repeats (light yellow) and palindromes P1–P8 (turquoise). **e. *Protein coding genes on chrY*.** Ubiquitously expressed single-copy genes (dark grey) reside in XDG and XTR regions whereas multi-copy genes (red) with mostly testes-specific expression are within AMP regions. The *TSPY* (testis-specific Y-encoded protein) gene family is arranged in a tandem array with 23–64 copies of a 20.4-Kb repeat unit containing a single copy of the *TSPY* gene, arrows show the gene directions. Figure adapted with permission from Box1 (Jobling and Tyler-Smith 2017), Springer Nature.

The five sets of inverted repeats (Figure 2 d.) are widely spaced in the AMP regions of chrY short and long arms, with repeat lengths of 62–298 Kb. Three of them (IR1, IR2, IR3) have nucleotide identity of 99.66–99.95%. The predominant long tandem arrays in the AMP regions are NORF (not annotated open reading frame) clusters with 622 Kb and a repeat unit of 2.48 Kb, and *TSPY* (testis-specific protein on Y chromosome) clusters that comprise 700 Kb and have a repeat unit of 20.4 Kb (Tyler-Smith *et al*. 1988). Due to high sequence similarity intrachromosomal homologous recombination and gene conversion is frequent in chrY and the ampliconic sequence is highly inclined to large-scale deletions, inversion and duplications that can have strong effect on the male fertility (Hughes and Rozen 2012; Rozen *et al*. 2003; Skaletsky *et al*. 2003).

Genes on Y chromosomes

The following gives a brief overview of the origins and functional categories of genes on chrY to highlight its biological importance. The reported numbers of genes and pseudogenes on chrY are at times different between studies depending on the annotation and validation methods as discussed in (Kuroki *et al*. 2006; Wilson Sayres and Makova 2013). A very simplistic example of this is given by a comparison of summary results from two public sources, Vega and Ensembl in Table 2 (Vega Genomic Browser 2017; Zerbino *et al*. 2017), reporting different numbers of protein coding genes on chrY.

Genes residing in different regions of human chrY can be: 1) genes with regular recombination between chrX and chrY in PAR regions; or in MSY region: 2) relics from the ancestral autosome (XDG region); 3) later additions to chrY from the chrX (XTR region); 4) multi-copy genes that are highly diverged amplified versions of the chrX-linked progenitors or derived and subsequently amplified from transpositions and retrotranspositions (AMP regions) (Figure 2 e.).

Across the MSY of different mammalian species highly non-random protein-coding gene sets with similar functions have been retained from the ancestral genes. The gene content of chrY reflects the retention, subsequent acquisition and amplification of gene families with testis-specific expression, and conservation of ancestral genes that need to be maintained on similar expression levels in males and females – the dosage-sensitive genes. The latter are ubiquitously expressed in the body and enriched for translation/transcription regulation and specific DNA binding/transcription factor activity (Bellott *et al*. 2014; Cortez *et al*. 2014). This leads to the proposal that in addition to the already appreciated roles of chrY in testis determination and sperm production through ampliconic gene families, chrY has a third organismal function – ensuring male viability through the single-copy genes (Bellott *et al*. 2014).

Table 2. Comparison of summary results for the annotations of human Y chromosome sequence and genes in two widely used public sources Ensembl (release 91) and Vega (release 68) (Vega Genomic Browser 2017; Zerbino *et al*. 2017)**.** NA – not analysed, lncRNA – long non-coding RNA)

chrY	**Vega 68**	**Ensembl 91**
Length (bp)	57,227,415	57,227,415
Finished Sequence	26,187,953	NA
Annotated Sequence	26,187,953	NA
Total Number of Clones	343	NA
Fully Annotated Clones	343	NA
Total Number of Genes	496	NA
Total Number of Protein Coding	48	63 coding genes
Known Protein Coding	48	NA
Total Number of Processed Transcripts	61	108 non-coding (NC) genes,
		30 small NC,
		70 long NC,
		8 misc NC
lncRNAs	59	NA
Unclassified Processed Transcripts	2	NA
Total Number of Pseudogenes	387	391
Processed Pseudogenes	136	NA
Unprocessed Pseudogenes	219	NA
Transcribed Processed Pseudogenes	4	NA
Transcribed Unprocessed Pseudogenes	28	NA
Short variants	NA	370,876

In the MSY region, genes from XDG and XTR regions are dosage-sensitive single-copy genes expressed ubiquitously in the body, except for the strict testes-specificity of *SRY*, the sex-determining gene from XDG region. The mostly or exclusively testis-specific multi-copy genes are expressed from AMP regions. These genes belong to nine MSY-specific gene families with > 98% sequence similarity between family members. The integrity of these genes specialized for male reproduction is retained through intrachromosomal gene conversion among the multiple family members on chrY. The majority of the transcribed genes in chrY are from the AMP region (Figure 2 e.) (Bellott *et al*. 2014; Bhowmick *et al*. 2007; Cortez *et al*. 2014; Delbridge 2004; Delbridge *et al*. 1999; Hughes and Rozen 2012; Lahn and Page 2000; Mazeyrat *et al*. 1999; Rozen *et al*. 2003; Saxena *et al*. 1996; Skaletsky *et al*. 2003).

2.2 Sequencing the Y chromosome

The male-specific inheritance and the lack of recombination in the large part of chrY makes studying the neutral variants of MSY a unique tool in population genetics. In essence, all the male lineages in the world are part of the global Y chromosomal phylogenetic tree. The geographic dispersal of chrY lineages together with their phylogenetic relations to each other carries information about migrations and demographic events in the male population; it also reflects the ancestral relations of the carriers of these paternal lineages.

Until the beginning of the second decade of 21st century the chrY phylogeny of contemporary males was studied by genotyping the limited number of single nucleotide variants (SNVs) diagnostic for particular branches on the tree. These markers were meticulously catalogued, their number was continuously growing, but often they were ascertained only in a small panel of individuals (Semino *et al*. 2000; Underhill *et al*. 2001). The systematic discovery of chrY polymorphisms globally was hindered because only limited resequencing was possible. The laborious iterative mapping of bacterial artificial chromosome (BAC) clones undertaken for the chrY reference sequence (Kuroda-Kawaguchi *et al*. 2001; Skaletsky *et al*. 2003; Tilford *et al*. 2001) was far too expensive for common use. In the lack of sequence data the time back to the most recent common ancestor (TMRCA) of contemporary male phylogeny was calculated mainly with information from the faster mutating short tandem repeats (STR), which can cause skewed time estimates (see below). Nevertheless, the approaches available uncovered a wealth of knowledge, but because of the limitations of data acquisition the genetic diversity of chrY was likely underestimated (Batini and Jobling 2017; Jobling, Mark A.; Tyler-Smith 2003; Jobling and Tyler-Smith 2017; Underhill and Kivisild 2007).

Since the first publication of the 1000 Genome Project Consortium (1000 Genomes Project Consortium *et al*. 2010) the availability of cost-effective short read sequencing (SRS, also known as the "next generation sequencing" or NGS) technologies have opened the cornucopia of genomic data and new horizons have opened also for chrY analyses (1000 Genomes Project Consortium *et al*. 2012; Batini *et al*. 2015; Drmanac *et al*. 2010; Fu *et al*. 2013, 2014; Hallast *et al*. 2015; Lupski *et al*. 2010; Poznik *et al*. 2013, 2016; Rasmussen *et al*. 2010; Rios *et al*. 2010; Roach *et al*. 2010; Tennessen *et al*. 2012; Wei, Ayub, Chen, *et al*. 2013). The SRS methodology produces from a randomly pieced sample genome a myriad of short (35–200 bp) sequencing reads that are mapped back to the reference genome sequence, a process reviewed in (Goodwin *et al*. 2016; Pfeifer 2016).

The mapping of these puzzle pieces back to the reference genome is achievable with high accuracy in regions where sequence content is unique or close to unique. The repetitive parts of the genome with low-complexity sequence, on the other hand, make it hard to assess exactly from which of these genomic regions a particular short read comes from. In these cases it becomes in fact impossible to uniquely map the short reads. Even when there are slight differences between sequences, it is hard to differentiate if the imperfect match to the one of the

repeats in the reference is due to the read from a nearly identical repeat or does the proband genome carry a true underlying variation relative to the reference. Sequencing errors add another layer of intricacy to this quest. Therefore, due to its complicated highly repetitive structure and regions with high homology to chrX, the sequencing of the "whole" chrY is technically impossible with SRS. Nevertheless, leaving the most repetitive parts aside, there are still substantial lengths of chrY to be sequenced (Figure 2 c.). Thus, more and more attainable SRS approaches are building a thriving database of chrY sequences, which, at the same time are a nidus for a new set of problems in the chrY data analyses.

Sequencing is characterized by coverage (the number of sites in the reference genome that have been covered by the sequence) and sequencing depth (number of reads covering each sequenced base). Both the depth and the total length of the assessed sequence are important parameters in discovering new variant nucleotides and giving more resolution for the underlying phylogenetic tree, the sequencing depth helps to distinguish between the true variants and false positives. There are several entry-points for the SRS of chrY, outputting data with varying depth and coverage. Some studies flow-sort the chrY prior to SRS sequencing (Xue *et al*. 2009), but mostly it is either chrY sequences from the complete human genome sequencing (Francalacci *et al*. 2013; Helgason *et al*. 2015; Poznik *et al*. 2013, 2016) or custom-based sequencing with baits designed to catch exclusively chrY regions (Barbieri *et al*. 2016; Batini *et al*. 2015; Hallast *et al*. 2015; Ilumäe *et al*. 2016; Lippold *et al*. 2014).

Besides differences in the physical length of chrY covered, also the sequencing depth varies across studies and needs to be taken into account, see (Poznik *et al*. 2016). Though constantly improving, the SRS technology has higher error rate compared to the earlier Sanger sequencing of ~0.1% vs 0.001% (Goodwin *et al*. 2016; Wang *et al*. 2012). The SRS sequencing results must be meticulously filtered to obtain final high quality sequencing calls. At the same time the high depth of sequencing provides a way around the false base-calling. When reads with sequencing errors make up just a tiny fraction of the overall read pool at a certain site the underlying base is called with high probability. The main cause of base calling errors is the poor mapping of the reads to the repetitive regions of the genome. Therefore, excluding low-complexity genomic regions from the mapping leads to much an improved sequence quality, but also reduces the overall coverage of the genomes that is attainable with this method. ChrY with its many low-complexity regions and single copy-number is uniquely the most challenging chromosome to be sequenced with SRS methods. It nevertheless offers also a practical approach to final high quality bases – the SNVs diagnostic for particular branches on the phylogenetic tree, proven to be trustworthy phylogenetic markers through the previous years of chrY study and mainly unique events in the human chrY tree. These polymorphisms can be used as a litmus test and reliable positive controls that provide helpful filtering criteria for the other calls (Xue *et al*. 2009).

Even when there is a wealth of high quality sequences from each individual the gaps or regions uncovered by sequencing reads are distributed unequally

among the samples. This results in the patchiness of the covered sequence in the whole dataset and needs special attention by finding the overlap between the well covered good quality regions in all the genomes under study. When merging datasets from different sources this common ground might in the end become considerably smaller than anticipated from original dataset sizes. Sometimes also imputing is used to overcome this problem (Barbieri *et al*. 2016). In addition, differences in SRS technologies cause another discrepancy, called *platform bias*. Each sequencing platform has its own unique error spectrum and that must be accounted for when merging data from distinct sources (Goodwin *et al*. 2016; Lam *et al*. 2012; Wall *et al*. 2014).

Therefore, even though the euchromatic male-specific region is about 24 Mb long, due to varying reasons, many parts of it are filtered out in the course of chrY SRS analyses. For example, ampliconic tracts for their high degree of self-identity, XTR region for the homology with chrX, and gaps in the reference sequence are most often excluded. The most frequently studied part of chrY, as of today, is the X-degenerate region (Figure 2 b., c.). Nevertheless, Helgason and colleagues detected *de novo* mutations in palindromic, ampliconic and XTR regions by combining the reads from paralogous positions and applying a weighting scheme to account for uncertainty about the location of the mutations (Helgason *et al*. 2015).

Regardless of these complications, piecing together the short reads eventually results in the amounts of chrY sequence never seen before, totalling 6–10 Mb of chrY sequence. It is crucial though, that any kind of variant detection and filtering would have similar criteria for variant and non-variant bases, so that not only variant positions, but the total length of the sequenced region can be estimated. Accurate estimation of the length of the sequence is important for the dating of evolutionary events since the underlying mutation rate is calculated as the amount of mutations accumulating through time in a sequence of certain length.

2.3 Mutation rate of the Y chromosome

Mutations fuel evolution and the dating of evolutionary events requires correct estimation of the mutation rates. The replication-associated mutations, the uncorrected errors in cell-division, are the most frequent type of alternations to genomes. They are caused by distinct molecular mechanisms with different rates that scale from a single base to whole chromosomes (Crow 2000; Wilson Sayres & Makova 2011; Jobling, Hurles 2014). In a multicellular organism only mutations in the germ line cells are hereditary and passed on to the next generation. In humans as in many other species the number of cell divisions in the germline prior to reproduction is higher in males and the difference increases with the fathers' age, a phenomenon called *male-driven evolution*. Because chrY is passed down only in the male germ line, its mutation rate is expected to be higher than the average of the whole genome (Haldane 1947; Kong *et al*. 2012; Li *et al*. 2002; Link *et al*. 2017; Wilson Sayres and Makova 2011).

SNVs and STRs are the two main types of genetic variation used to characterize chrY evolution. SNVs are single genomic point loci, nucleotides that differ between individuals, whereas STRs are highly polymorphic and mutable loci with 2–6 bp long repeat units spanning a median of 25 bp. There are approximately 700,000 STRs occupying in total ~1% of the human genome (Willems *et al*. 2014). About 4,500 chrY STRs have been examined in the SRS data, the majority of these were fixed in the population and ~700 displayed length polymorphism (Willems *et al*. 2016). STR sites have high mutation rate because in the replication process the repetitive nature of the underlying DNA template causes the polymerase to slip, leading to the deletion or addition of the repeat motif(s) (Ellegren 2004; Weber and Wong 1993). The fast mutation rate of chrY STRs provides a high discriminative power between males and therefore they are widely used in forensics, population genetic and genetic genealogy studies. The general rule is that the larger the number of STRs that individuals are tested for, the higher the discriminative power (Kayser 2017), which is especially important in forensics. Still there are sets of standard chrY STRs that are very widely used for their robust results, cost-benefit ratio, and applicability in different study questions which all provide comparability between studies. The sets consist of 17 STRs in the Y-filer kit of Applied Biosystems and 23 STRs (including all previous 17) in the PowerPlexY23 of Promega. These have been the main work horses in the study of STR diversity of chrY, even though many more chrY STRs are characterized (Ballantyne *et al*. 2010; Burgarella and Navascués 2011; Willems *et al*. 2016).

A comprehensive review of the current state of the art in chrY mutation rate studies is given in (Balanovsky 2017), overviewed here only briefly. There are three main approaches for estimating the chrY mutation rate: the *direct genealogical* method of counting mutations along the genealogy, the *calibration* approach utilizing a dated event in a population's history, and the *ancient chrY* (aY) approach of counting the missing mutations when compared to modern lineages (Figure 3.a.). The calibration and the aY approaches are collectively also called *evolutionary rate estimations*. These approaches are applied both for SNVs and STR rates, except for the aY, which is used for SNVs only. In addition, human chrY mutation rate has been estimated from the interspecies comparison and by correcting the autosomal mutation rate for chrY. The first method utilizes the interspecies difference in numbers of chrY mutations in relation to the divergence time between human and chimpanzee (~6 MYA) (Kuroki *et al*. 2006), whereas the other corrects the previously calculated genealogical autosomal mutation rate (Kong *et al*. 2012) for the fathers' age at conception to reach the chrY mutation rate (Mendez *et al*. 2013).

The *direct genealogical* methods measure mutation rate from the individual meiosis in the pedigrees or in a single generation between parents and children (Figure 3 a.1). Well documented deep-rooting pedigrees that trace the male lineage over several generations to the common forefather with known birth year provide the possibility to study more meiosis between a limited number of

individuals, giving a more accurate rate estimate (Balanovsky *et al*. 2015; Helgason *et al*. 2015; Heyer *et al*. 1997; Xue *et al*. 2009) (Table 3).

Calibration methods (Figure 3 a.2.) use some external data, either well-characterized and dated fossils, in cases of species' divergence together with species' genetic divergence data (Kuroki *et al*. 2006; Thomson *et al*. 2000) or well-documented colonization events in setting the upper bound to the coalescence of lineages (Poznik *et al*. 2016; Zhivotovsky *et al*. 2004).

The *aY method* introduced in (Rasmussen *et al*. 2010) calculates the 'number of missing mutations' (Figure 3 a.3.) in the ancient sample compared to its closest relatives among the currently living individuals and calibrates this number with the archaeological age of the ancient sample. The number of missing mutations is proportional to the age of the sample; the older the date, the less time there was to accumulate mutations. To date, the narrowest confidence intervals for this rate calculation method come from the aY with highest coverage from a Siberian Palaeolithic sample (Fu *et al*. 2014) (Table 3).

Increased availability of large scale chrY and aY sequence data, with up to ~10 Mb of sequence and tens of thousands of assessed variants, have fuelled several studies on SNV mutation rate (Balanovsky *et al*. 2015; Francalacci *et al*. 2013; Fu *et al*. 2014; Helgason *et al*. 2015; Karmin *et al*. 2015; Poznik *et al*. 2013; Trombetta, D'Atanasio, Massaia, Myres, *et al*. 2015; Xue *et al*. 2009). Their main results are summarized in Table 3 and Figure 3.

When sequencing data was scarce the chrY STRs diversity was the main tool to study recent paternal demographics (Batini and Jobling 2017). The STR mutation rate has been estimated mostly from the pedigrees and father-son pairs in numerous studies and 29 of these were meta-analysed in (Burgarella and Navascués 2011).

The STR mutation rates differ significantly between the loci and the rate calculation is dependent on the underlying mutation models (Ballantyne *et al*. 2010; Burgarella and Navascués 2011; Heyer *et al*. 1997; Kayser *et al*. 2000; Willems *et al*. 2016). The pedigree-based rates converted to the per locus per year rate (with generation time of 31 years) for the most widely used sets of STRs in Y-filer and PowerPlexY23 are ~8.6×10^{-5} and ~12.5×10^{-5} per locus per year respectively (Balanovsky 2017). These rates result from many studies (Ballantyne *et al*. 2010; Burgarella and Navascués 2011; Busby *et al*. 2012; Järve *et al*. 2009; Ravid-Amir and Rosset 2010; Zhivotovsky *et al*. 2006) with increasing precision, see Table 2 in (Balanovsky 2017). The single evolutionary mutation rate estimate for chrY STRs that has been widely used in population genetic studies was calculated as the average of three different STR mutation rate estimates, resulting in the rate of 2.8×10^{-5} per locus per year (Zhivotovsky *et al*. 2004). The three-fold difference between the genealogical and evolutionary STR mutation rate has been explained by the effect of recurrent mutations and genetic drift which saturates the STR diversity through time (Zhivatovsky 2006, Wei 2013a,b, Hallast 2014). The difference between the genealogical and evolutionary mutation rate for the SNVs is also present, but in much smaller scale than that of STRs (Table 3, Figure 3b).

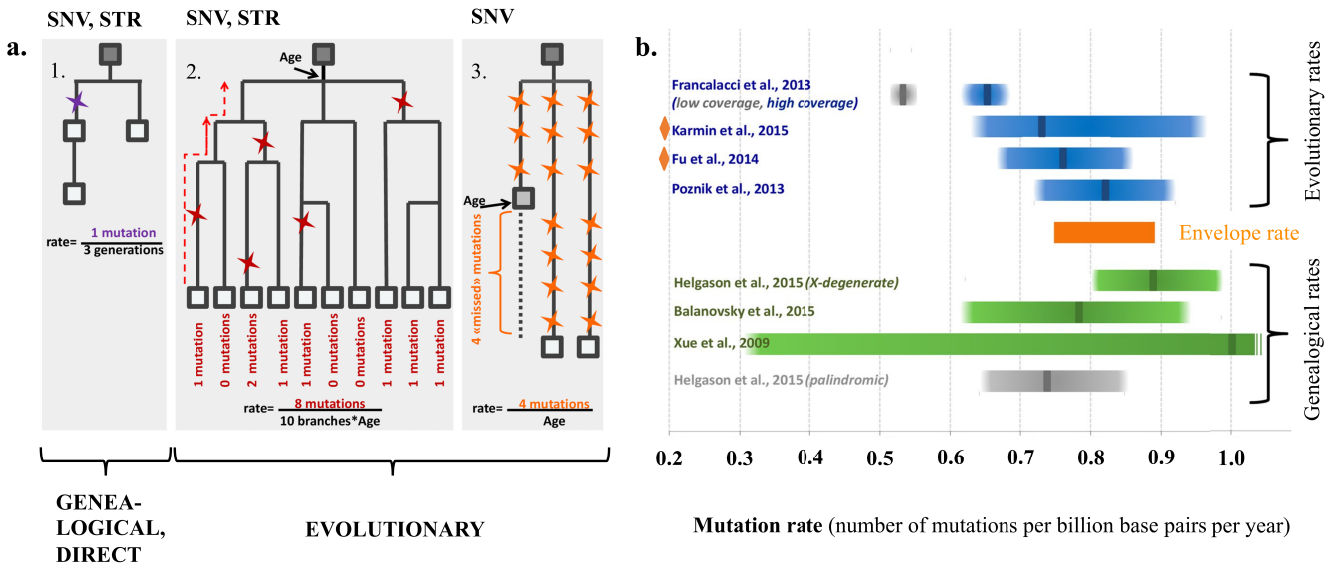

Figure 3. Three approaches of chrY mutation rate estimation (a.) and the chrY SNV mutation rates (b.). a.1 The *direct estimation* of mutations: mutational differences between the members of the pedigree are counted and divided by the number of generations in the genealogy. **a.2**. The *calibration method*: the number of mutations between the MRCA and the contemporary samples is averaged and divided by the TMRCA, based on the assumption that TMRCA coincides with a dated population event. **a.3.** *Ancient Y method*: a chrY from a dated ancient specimen is missing some mutations compared to the contemporary individuals because its evolution stopped at the point of time when the individual died. The number of the missing mutations is proportional to the age of the sample, the older the date the less time the sample has had to accumulate. Dark squares show ancestral individuals/lineages; white squares show contemporary individuals; light grey square show ancient individual, stars denote mutations. **b. The SNV mutation rates from recent chrY sequencing studies.** Rate in number of mutations per billion base pairs per year (x-axes). Point estimates shown as darker bars, confidence intervals in lighter colour. Blue – evolutionary rates; green – pedigree rates; the orange line is the 'envelope rate' used in (Poznik *et al*. 2016). Grey – estimation from low coverage sequencing data (Francalacci *et al*. 2013) and from the palindromic regions of chrY (Helgason *et al*. 2015). Orange diamonds denote the aY estimations. Figure adapted with permission from Fig.1 and 3 of (Balanovsky 2017), Springer Nature, Human Genetics.

Table 3. The SNV mutation rates from Y chromosome short read sequencing studies.
Adapted with permission from Table 1 of (Balanovsky 2017), Springer Nature, Human Genetics.

Study	App-roach	Rate[a]	95% CI ($\times 10^{-9}$)	Sequenced length (Mb)	Cove-rage[b]	Note
Xue et al. (2009)	Genealogical	1×10^{-9}	0.3–2.5	10.15	>11	Deep-rooting paternal pedigree, 2 individuals, 13 meioses.
Helgason et al. (2015) *X-degenerated region*		0.89×10^{-9}	0.80–0.99	8.96	>10	Deep-rooting paternal pedigree, 753 individuals, 1365 meioses.
Helgason et al. (2015) *Palindromic region*		0.74×10^{-9}	0.64–0.85	6.24	>10	
Balanovsky et al. (2015)		0.78×10^{-9}	0.62–0.94	9.97	>45	Nine persons from the same clan with the forefather with known birth date of 1405 CE.
Poznik et al. (2013)	Calibration	0.82×10^{-9}	0.72–0.92	9.99	>5	Calibration based on first human colonization of Americas and expansion ~15 KYA taken as the the split of Q-M3 lineages from Q-L54*(xM3).
Francalacci et al. (2013) *low coverage*		0.53×10^{-9}	0.52–0.55	8.97	>1	Calibration based on the Sardinian specific sublineage I2a1a-δ and the peopleing of Sardinia ~7.7 KYA.
Francalacci et al. (2013) *deep coverage*		0.65×10^{-9}	0.62–0.68	8.97	>14	
Fu et al. (2014)	Ancient DNA	0.76×10^{-9}	0.67–0.86	1.86	>22	~45KY old Ust'-Ishim aY (hg NO) and 23 contemporary hg NO individuals.
Karmin et al. (2015)		0.74×10^{-9}	0.63–0.95	8.82	>4	~12.6 KY old Anzick (hg Q1b) and ~4 KY old Saqqaq (hg Q2b) aY with contemporary sameples (17 and 5 respectively). Tranversions only.
Trombetta et al. (2015)		0.72×10^{-9}	0.62–0.82	1.5	>2	~45 KY old Ust'-Ishim aY (hg NO) and a ~7 KY old European Loshbour (hg I) aY together with 104 contemporary individuals.

[a] Average number of mutations per base pair per year
[b] Of that sample in the dataset which has the lowest coverage
CE – current era
KY(A) – kilo years (ago

The obvious question as to which rate to use, is both practical and that of curiosity. For SNVs the suggestion is to use an "envelope" rate that is based on the average rate from aDNA-based and that of the pedigree rate (Poznik *et al*. 2016). The estimated envelope rate for the high-coverage sequencing of XDG is 0.75–0.89 substitutions per billion base pairs per year (Balanovsky 2017; Poznik *et al*. 2016). The comparison of STR and SNV rates, (Wei, Ayub, Xue, *et al*. 2013) shows that STR-rates underestimate the age of the whole chrY tree. Hallast and colleagues (Hallast *et al*. 2015) demonstrate that the 'evolutionary' STR rate is in a better accordance with the age estimates of older events, whereas the genealogical rate suits better for the younger lineages (Balanovsky 2017).

It is important to note that mutation rate estimates come within a range of uncertainty stemming from the dating of fossils, sequencing errors, missing data, limited number of observed meiosis and the underlying uncertainty of the mutation process itself. Therefore, improved data for any of these factors would provide a more precise chrY mutation rate estimates.

In the human population history and demographic reconstructions the mutation rates are applied to date coalescent times in the phylogenetic tree. One widely used method for this is the *rho*-statistics that takes into account the average number of differing sites between a set of sequences and a specified common ancestor (Forster *et al*. 1996; Saillard *et al*. 2000). Another popular age calculation method is the coalescent-based Bayesian estimation, implemented in the software BEAST (Drummond and Rambaut 2007). Dates from the coalescent-based method and *rho*-statistics correlate strongly in case of SRS data (Batini *et al*. 2015; Scozzari *et al*. 2014; Trombetta *et al*. 2015).

2.4 Male *vs* female genetic variation in human populations

The male genetic diversity between populations has been reported to be higher than that of females. This manifests in higher between-group F_{ST} values for chrY, compared to those of mtDNA, arguably due to the patrilocality of populations (Seielstad *et al*. 1998; Stoneking 1998). Though this finding has been attributed also to methodological differences in assessing the variation in mtDNA and crhY during earlier studies (Wilder *et al*. 2004), more recent analyses of sequence data for both chrY and mtDNA show similar results. Differences in the relative amount of between-populations vs within-populations genetic diversity for these loci are demonstrated, but with a smaller scale than previously reported and with large differences in regions of the world (Lippold *et al*. 2014).

Indeed, the residence pattern of the population – matrilocality or patrilocality – has been used to explain this phenomenon on a local geographic range (Gunnarsdóttir *et al*. 2011; Oota *et al*. 2001; Ségurel *et al*. 2008; Wilkins and Marlowe 2006). Typically, higher genetic differentiation between groups for mtDNA in matrilocal populations has been observed, whereas the opposite is true for

patrilocal populations. Exceptions to this rule have included the Mon Khmer in Thailand (Besaggio *et al*. 2007) and some Indian matrilocal populations (Kumar *et al*. 2006), highlighting the complexity of the issue when other cultural practices override the residence rule as the main influence on the differential genetic diversity of males and females. Importantly, when geographically more distant populations are compared, the effect of the residence rule declines and differences of F_{ST} values for mtDNA and chrY decrease. The longer physical distance between populations usually also means that the shared history is more distant in time.

The decline in the relative differences of female and male genetic diversity between populations geographically far apart could be explained by the more widespread matrilocality in the distant past. The ancient signal obliterates the more recent local effects of patrilocality on the genetic diversity. It has been shown that among the Bantu and Bantoid-speaking populations in sub-Saharan Africa matrilocality was widespread before the current custom of patrilocality, which was established together with the onset of animal husbandry (Holden and Mace 2003). This gives way to the possibility that the Neolithic period could be the time when patrilocality started to dominate. However, the matter is complicated by the fact that the sequence of the residence rule throughout the history of humankind is not clear (Copeland *et al*. 2011; Holden and Mace 2003; Jordan *et al*. 2009; Lalueza-Fox *et al*. 2011), just as it is possible that the rules might have changed several times.

At the same time, the genetic variation patterns of a population are strongly influenced by the number of mating individuals that contribute to the gene pool of the population, the effective population size (N_e). The effect of genetic drift is dependent on the N_e, the smaller the number of individuals contributing to the population's gene pool, the stronger the drift, which, in turn, leads to lower diversity within the population and higher level of differentiation between populations as they drift apart. Therefore, the observed higher chrY diversity between populations, compared to that of mtDNA, could be the consequence of smaller N_e of men (Heyer *et al*. 2012). Thus, besides the sex-specific behaviours – the proposed higher migration rates of women in more recent times, the higher differences of chrY variation between populations can be due to stronger drift acting on this locus. In essence, the cumulative effect of migration and N_e together influence the strength of the drift (Chiaroni *et al*. 2009).

To study the influence of the spread of Neolithic and agriculture from the Near East to Europe on the genetic variation of European population the mtDNA and NRY haplotype frequency data was analysed jointly within the same framework. Demographic events were modelled with a simple admixture model (Figure 4, right panel) (Rasteiro and Chikhi 2013). Both datasets supported the demic diffusion model of agriculture to Europe, but also some sex-specific differences became apparent.

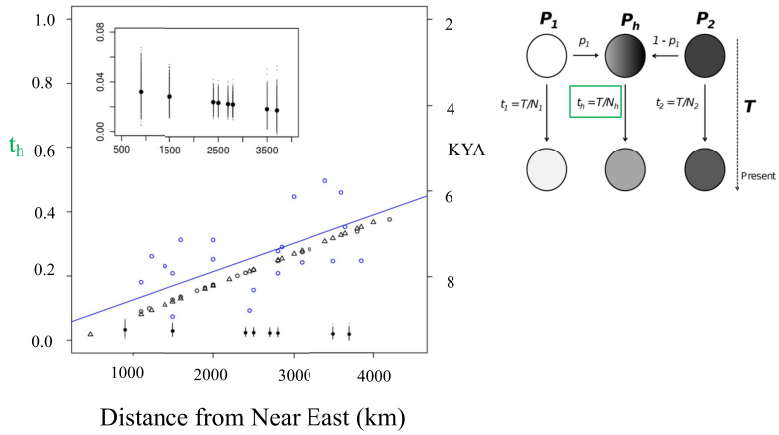

Figure 4. The longer time spent in small structured groups increases the influence of drift in European male populations as the geographic distance from the Near East grows. Linear regression of drift parameter t_h (y-axes on the left) for mtDNA and chrY data of contemporary European populations plotted against the geographic distance from the Near East (x-axes). Drift since the admixture event between the Neolithic farmers from the Near East and local hunter-gatherers, modelled as shown on the right panel, is dependent on the time and effective population size ($N_1/N_h/N_2$). According to the model an admixture event with a fixed amount of genetic contribution from Palaeolithic local hunter-gatherers and Neolithic farmers occurs, then ancestral and admixed populations all are influenced by the drift (t_i) alone. P_1 – Palaeolithic population (Basques), P_2 – Neolithic population (Near Eastern populations pooled for mtDNA; Turks for chrY), P_h – admixed population (contemporary European populations), p_1 – the Palaeolithic genetic contribution, (1-p_1) – Neolithic genetic contribution, T – time since admixture event. N_1 – effective size of the Palaeolithic population, N_h – effective size of the admixed population, N_2 – effective size of the Neolithic population. Left panel: empty triangles and circles – chrY datasets (Rosser *et al.*, 2000; Semino *et al.*, 2000 respectively), filled circles – mtDNA dataset (Richards *et al.* 2000). Blue empty circles – calibrated radiocarbon dates from European Neolithic archaeological sites (timescale on the Y-axes on the right) plotted as distance from the Near East, blue line as a linear regression for this set. The chrY data shows clear increase of drift as the distance from the Near East grows, whereas for mtDNA such trend is not visible (see inlay for close-up). For chrY drift is stronger in populations that have been small structured hunter-gatherer groups for longer time; the global effective population size for males is larger in the Near East. Female population has had a larger effective size and/or more mobility between groups. Figure adapted from Figure 1 and Figure S1 (Rasteiro and Chikhi 2013), licensed under the CC-BY License.

The linear regression shows that male populations have a strong positive correlation between the genetic drift parameter and the geographic distance from the Near East, whereas for females such correlation is absent (Figure 4). For chrY, the drift is stronger in populations that are further away from the Near East and have been for longer time hunter-gatherers, thus spent more time in structured populations and had smaller effective population size of males (N_m), making chrY diversity more prone to drift. At the same time the global N_m is larger in the Near East and decreases with distance from the Near East. The lack of such correlation

in the female population (Figure 4) points to the larger effective population size of females (N_f) and/or more mobility of females than males between groups.

The skyline plot method (Ho and Shapiro 2011; Pybus *et al*. 2000), implemented as a Bayesian skyline plot (BSP) technique in the software BEAST (Drummond *et al*. 2005; Drummond and Rambaut 2007), summarizes the dynamics of N_m and N_f through time from the DNA sequence data. Comparing mtDNA complete sequences and 0.5 Mb of chrY sequence from 623 males from 51 populations confirmed the previously reported larger genetic differences for male lineages between global populations, but on a smaller scale (Lippold *et al*. 2014). Model-based simulations showed extremely low (<100) ancestral N_e during the OOA bottleneck (Figure 5). This study also revealed that N_f has been larger throughout human history and starting at ~15–20 KYA N_f had a larger growth rate than N_m (Lippold *et al*. 2014).

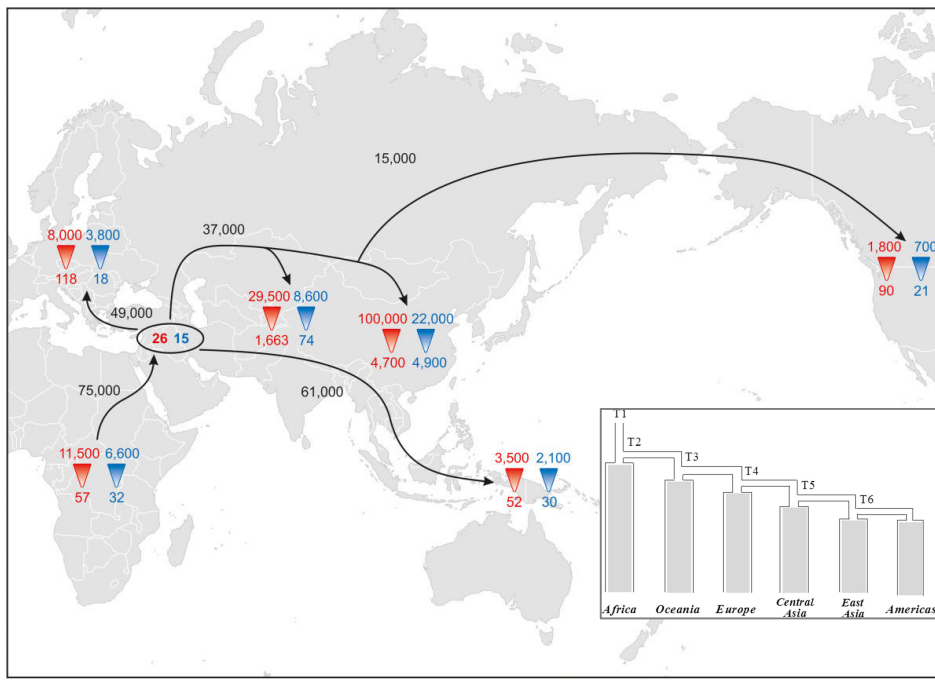

Figure 5. Global female (N_f) and male (N_m) effective populations sizes (N_e) and divergence times. N_f (red triangles) and N_m (blue triangles) and divergence times (black numbers) estimated with the Approximate Bayesian Computation method following the model given in the inset. The data is simulated from complete mtDNA and 0.5 Mb chrY sequences from 623 individuals from Human Genome Diversity Panel. The ancestral N_e is noted at the bottom of the triangles and current at the top. N_f – red text and triangles, N_m – blue text and triangles. The number in the oval denotes the Out of Africa population bottleneck and estimated number of individuals for N_f and N_m. The arrows indicate the direction of migrations and do not depict the actual migration routes. Figure adapted from Figure 1 and Figure 6 (Lippold *et al*. 2014) licensed under the CC-BY License.

Abrupt changes in the male gene pool were reported also by the population based chrY resequencing of 1,224 individuals from 26 global populations (Poznik *et al*. 2016) and from more targeted 334 samples of European and Middle Eastern descent (Batini *et al*. 2015). A picture of recent, rapid and continent-wide demographic events that have had major impact on the current European paternal gene pool emerged. Over half of present European paternal lineages stem from three founders coalescing no more than 7.3 KYA (Batini *et al*. 2015) and lineages belonging to R1b-L11 expanded around 4.8–5.5 KYA (Poznik *et al*. 2016). These events are suggested to be the signs of the rapid changes also reflected in the archaeological culture of the Bronze Age during the time of massive migrations towards Europe (Allentoft *et al*. 2015; Haak *et al*. 2015).

Recently, a study modelled two or more male populations competing with each other and mating with the same population of females. The simultaneous tracking of the dynamics of cultural groups and chrY hgs, showed that the decrease of N_m (with more or less constant N_f) can be the result of patrilineal kin groups with intergroup competition among them (Zeng *et al*. 2018). Thus, cultural and social practices, such as the change of lifestyle and ways of subsistence, sedentism, and residence rules can have a strong influence on human genetic diversity patterns. The genetic diversity patterns of human male (and female) gene pool are shaped by a complex interplay between the demographic, evolutionary and cultural forces.

2.5 Phylogeny and phylogeographic distribution of Y chromosome haplogroups

Background of phylogenetic studies, nomenclature

The sequential accumulation of mutations in the MSY makes chrY lineages traceable in a phylogenetic tree. Within a phylogenetic tree the lineages that are related to each other more closely than to anyone else are shown as nested monophyletic groups (clades), characterized by specific common ('diagnostic') variants and called 'haplogroups' (hg). In 2002 Y Chromosome Consortium (YCC) unified the naming of hgs of the rooted chrY phylogenetic tree. Almost all the known chrY SNVs at the time (237) were used for defining 153 haplogroups with an alternating alphanumeric nomenclature. The combined system labelled horizontally 18 primary branches by capital letters (A to R) and vertically each clade was assigned a unique alphanumeric label while unlabelled clades could be named as 'join' of the sub-clades. The vertical nested scheme allowed precise mapping of clades on the tree. The term 'paragroup' was coined to denote lineages that represent interior nodes of a haplogroup, but are not (yet) defined on the basis of derived state in any known DNA-variant and are genetically paraphyletic. Paragroups are marked with the asterisk (*) symbol, representing chromosomes belonging to a clade but not to any of its (known) subclades. At the same time an alternative naming scheme was proposed – the

major haplogroup name (i.e., 18 capital letters) followed by the name of the terminal variant that defines a given haplogroup (The Y Chromosome Consortium 2002). Later important updates to the original YCC tree were published, refining the tree to have 20 main haplogroups (A to T), but encompassing still some multifurcations among the major splits of the tree (Jobling and Tyler-Smith 2003; Karafet *et al*. 2008).

This established nomenclature was fruitfully applied in many chrY studies that in turn also revealed thousands of new chrY diagnostic variants. The exact phylogenetic placement and diagnostic value of new markers had to be disclosed by laborious genotyping in big cohorts. The information was later carefully catalogued by ISOGG (https://isogg.org/). Due to higher availability of finances, samples and public interest in populations of Western European descent, for long, the chrY variants in other populations were poorly characterized. Therefore, the possibility to sample continuous chrY sequences from diverse haplogroups and populations was long waited for. The shift in data acquisition uncovers chrY variants systematically, so the most informative ones for particular lineages can be chosen for further genotyping in larger cohorts.

The explosion of chrY variants pouring in from the SRS studies (Batini *et al*. 2015; Francalacci *et al*. 2013; Hallast *et al*. 2015; Mendez *et al*. 2013; Poznik *et al*. 2013, 2016; Wei, Ayub, Chen, *et al*. 2013; Xue *et al*. 2009) provided an ever-higher resolution to the phylogeny, but also demanded constantly more effort to catalogue all the variants to the point where the choice of the most informative diagnostic genotyping markers became increasingly more complicated. Meanwhile a 'minimal reference tree' was published with the aim to have an optimal global discrimination capacity based on a strongly reduced set of diagnostic variants (417 at the time of publication). These include only the most resolving Y-SNVs with common standards for Y-marker as well as Y-haplogroup nomenclature. The authors' main aim was to provide a state-of-the-art phylogeny to include all the changes made since the publication of previous reference phylogeny, but also to retain as much as possible of the historical hg labels and marker names for consistency through studies (Van Oven *et al*. 2014).

Main rearrangements of the topology and hg labels were in the deepest parts of the tree, following the proposals from several earlier studies (Batini *et al*. 2011; Cruciani *et al*. 2011; Mendez *et al*. 2013; Van Oven *et al*. 2014; Scozzari *et al*. 2012). In the influx of new variants that revealed more of the tree structure, the hg labels are often kept with their initial markers and new (joint) labels are given to newly resolved structures, i.e., hg label K has been kept with marker M9 and K2 defined with M256 (Poznik *et al*. 2016). This illustrates the difficulty of keeping up an all-comprehensive nomenclature with the current amount of new data. The system of using the 'defining marker' names is more robust, but comes with the drawback of many variants having several pseudonyms.

Topology and contemporary phylogeography of major Y chromosome haplogroups

The deepest splits and the deep African root of the Y chromosome phylogeny

It has been long established that both maternal (Behar *et al*. 2008; Cann *et al*. 1987) and paternal lineages stemming from the deepest splits of the tree are exclusively African. This supports the model according to which modern humans arose in Africa and the rest of the world was colonized by a subset of the initial African population (Cruciani *et al*. 2011; Hallast *et al*. 2015; Hammer *et al*. 1998; Poznik *et al*. 2016; Semino *et al*. 2002; Stringer 2002; Tishkoff *et al*. 2007; Underhill *et al*. 2001).

Sequencing chrY in samples from specific hgs has profoundly refined the deeper part of the phylogeny, revealing five deep splits that define hgs **A00, A0, A1, A2, A3** and **B.** These hgs encompass not so frequent or extremely rare, but exclusively African lineages (Barbieri *et al*. 2016; Batini *et al*. 2011; Cruciani *et al*. 2011; Mendez *et al*. 2013; Poznik *et al*. 2013; Scozzari *et al*. 2012, 2014). The longest internal branches within the tree are among those hgs and some of these lineages, different sub-lineages of hg A and B2b, have been mainly observed among hunter-gatherers, representing ancient paternal lineages (Cruciani *et al*. 2002; Underhill *et al*. 2000). Other nested clades like B2a are among those proposed to be the paternal signatures of the extensive and more recent population movements related to the spread of agriculture, iron work and Bantu languages within Africa (Beleza *et al*. 2005; Cruciani *et al*. 2002; Scozzari *et al*. 2014; Underhill *et al*. 2000).

TMCRA of global human Y chromosome phylogeny

For long, the estimates of the TMRCA of chrY were considerably younger, even twice as young as that of mtDNA. The coalescence time of chrY ranged within 50–115 KYA (Pritchard *et al*. 1999; Tang *et al*. 2002; Thomson *et al*. 2000), with few older estimates (Cruciani *et al*. 2011; Hammer 1995). This discrepancy was mainly caused by the lack of sequence data and SNVs for correct calibration.

Analyses based on ~10 Mb of chrY sequence derived from 69 men, including those carrying a diverse set of hg A lineages, suggested that human chrY tree coalesces at ~139 (120–156) KYA (Poznik *et al*. 2013). This estimate appears to be in the comparable range with the mtDNA phylogeny of the same individuals, coalescing ~124 (99–148) KYA. However the concurrent discovery of the deepest known chrY lineage – A00, from an African American genetic genealogy customer, shifted the coalescence age estimate for the global human chrY tree drastically to ~338 (237–581) KYA (Mendez *et al*. 2013). Thus far, hg A00 has been found only among the Mbo-speakers from Cameroon and not among extant nor ancient hunter-gatherers from sub-Saharan Africa (Barbieri *et al*. 2016; Schlebusch *et al*. 2017; Skoglund *et al*. 2017). Subsequent re-analyses of this lineage and the Neanderthal El Sidrón chrY revealed that the global human chrY tree coalesces ~275 (243–304) KYA and splits from the Neanderthal lineage at about ~588 (447–806) KYA (Mendez *et al*. 2016). The

respective differences in the coalescence ages of the global chrY tree result from the methods the TMRCA was estimated. The coalescence time between A00 lineages and the rest of the chrY tree was initially estimated using the autosomal mutation rate adjusted for male-specific mutation processes (Mendez *et al*. 2013). The later study (Mendez *et al*. 2016) used the known age of ~45 ky for the aY of Ust-Ishim sample (Fu *et al*. 2014) in the Poisson process to model the mutations on the tree and calculate the maximum likelihood TMRC estimate together with the mutation rate as described in (Rasmussen *et al*. 2014).

The global phylogeographic distribution of most common paternal haplogroups

The monophyletic cluster CT-M168 (aka CDEF) that splits ~76,000 years ago (YA) into sub-clades DE-M145 and CF-P143 (Figure 6) encompasses all other haplogroups that today form the majority of the African chrY lineages and all of the non-African lineages (Hallast *et al*. 2015; Poznik *et al*. 2013, 2016; Underhill and Kivisild 2007).

All lineages downstream from CF-P143 are spread exclusively outside Africa, but the DE-M145 holds hg **E-M96**, the current predominant African hg that also has a marked spread outside the continent in the Mediterranean Europe (Figure 6 and Figure 7) (Semino *et al*. 2004). Its sister lineage hg D-M147 has a confined East Asian distribution. The sub-lineages of **hg D** have a fragmented and population-specific spread, being at higher frequency among Tibetans and Japanese and some non-Tibetan Tibeto-Burman speaking populations, but also found among the Y chromosomes of the Andamanese islanders (Figure 6 and Figure 7) (Chiaroni *et al*. 2009; Poznik *et al*. 2016; Shi *et al*. 2008; Thangaraj *et al*. 2003).

The distribution and phylogenetic relationship of D and E lineages has fuelled a debate about the birthplace of their parent clade. The wide African distribution and diversity suggested the East Africa as the place of origin of hg DE and the subsequent migration of hg D out of Africa (Cruciani *et al*. 2004; Underhill *et al*. 2001; Underhill and Kivisild 2007). The most parsimonious phylogenetic model with only one continental exchange proposes that hg DE arose outside Africa with back migration of hg E to Africa, instead of hgs D, C and F moving out of the continent (Poznik *et al*. 2016) (Hammer *et al*. 1998).

The main non-African clade CF-P143 encompasses all other clades of the global chrY phylogeny with distribution outside of sub-Saharan Africa. It splits to hg C-M130 and megahaplogroup F-M89 (Figure 6). **C-M130** has a high to moderate frequency in Far East, Oceania and Australia and the northern part of North America. Its proposed origin is in the northeast Asia (Figure 7). Its specific sub-lineages are present in populations of Central-Asia, Mongolia and Siberia (C3-M217) with further sub-lineages among Native Americans (C3-P39), Pacific Islanders and populations in the highlands of Papua New Guinea (C2-M38). Another sub-branch (C4-M347) is the most common hg of Australians, whereas C5-M356 in Pakistan, India and Nepal (Bergström *et al*. 2016; Gayden *et al*. 2007; Hammer *et al*. 2006; Kayser *et al*. 2006; Sengupta *et al*. 2006; Zhong *et al*. 2010).

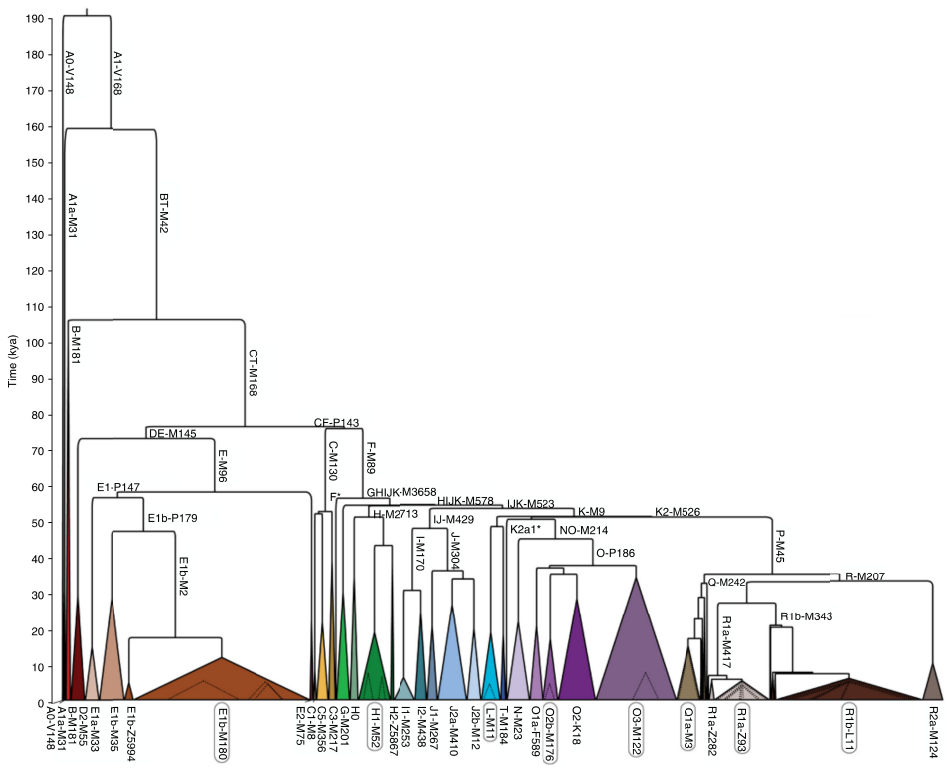

Figure 6. A calibrated Y chromosome (chrY) phylogenetic tree. The phylogenetic tree based on 60,555 SNVs from 10 Mb of chrY sequence of 1,244 individuals from 1000 Genomes Project. Tree is calibrated with the mutation rate of 7.6×10^{-10} mutations per base pair per year (Fu *et al*. 2014). Haplogroup (hg) names are below the triangles. The hgs with many chromosomes are collapsed into triangles, height represents coalescence time and width is proportional to the number of individuals in the clade. Grey ovals mark the hgs showing rapid recent expansion. Two major lineages, both subgroups of hg K2b1 – hgs M and S, mostly prevalent in the Island South East Asia and New Guinea, are absent. Figure adapted with permission from (Poznik *et al*. 2016), Springer Nature.

The **megahaplogroup F-M89** is the other branch splitting form CF-P143. A notable increase in the expansion of lineages outside Africa around ~50–55 KYA occur within the F-M89 clade (Figure 6 and Figure 7). The megahaplogroup comprises all lineages from the GHIJK clades, but first split within F-M89 is the newly discovered F* branch found in a Vietnamese individual (Figure 6) (Poznik *et al*. 2016).

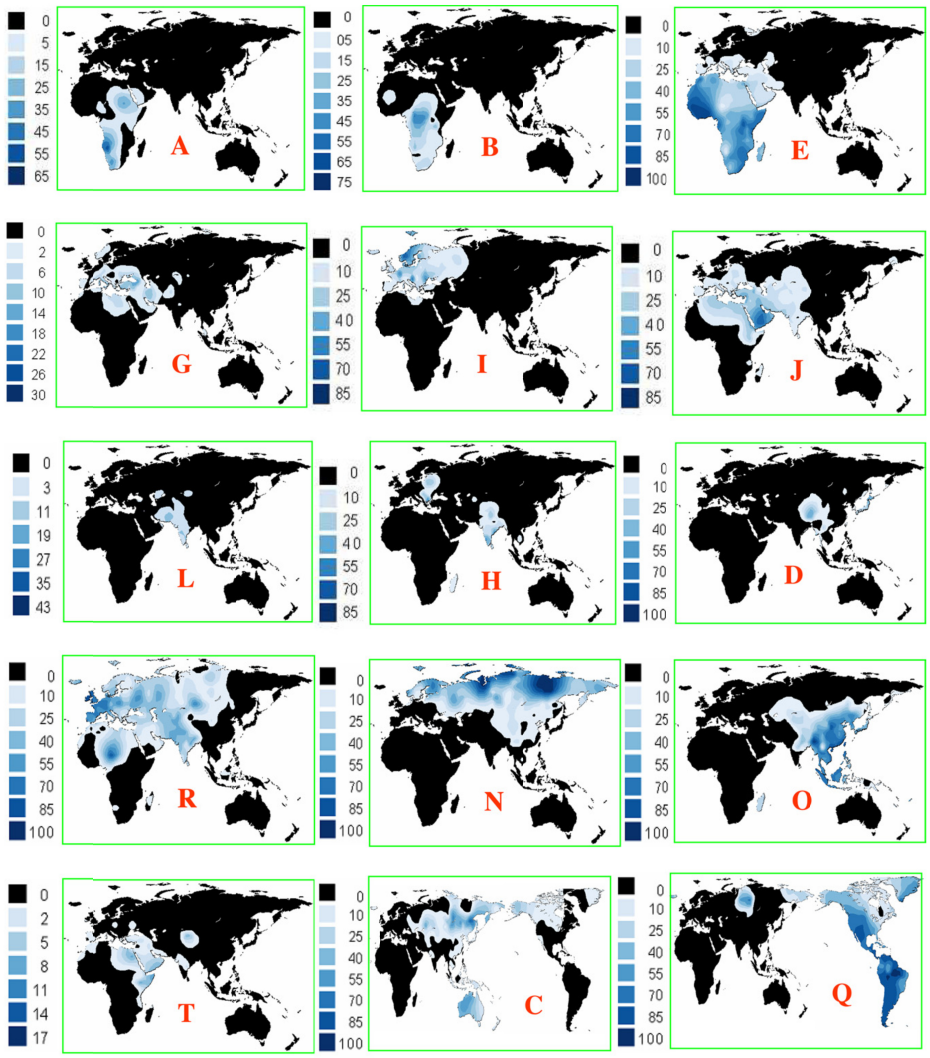

Figure 7. The global phylogeographic distributions of 15 numerically and geographically most widespread chrY haplogroups. The distributions of hgs K, M, and S are not displayed because of their very distinct special distributions, mostly in the Island South East Asia and Oceania. Also, paraphyletic hgs F* and P* are mostly rare, and not shown. Hg F* shares the distribution with hg H within India. Figure reprinted with permission from (Chiaroni *et al*. 2009).

Haplogroup **G-M201** (Figure 6, Figure 7) has highest frequency in the Caucasus and spreads to the Near and Middle East, parts of northern Africa and southern Europe. The frequency peaks among North Ossetians (70%) and declines (to 13%) in Iran, but the hg is found also in the South Asian populations (Regueiro *et al*. 2006; Rootsi *et al*. 2012; Semino *et al*. 2000; Sengupta

et al. 2006; Yunusbayev *et al*. 2012) and northern Africa (El-Sibai *et al*. 2009). It has two phylogenetic lineages G1-M285 and G2-P287 (Figure 6). G1-M285 has a relatively low frequency in its main distribution area in South-West Eurasia, where it ranges from Italy to Mongolia. The sub-haplogroup reaches highest frequencies in Central Asian steppes and Iranian-Armenian plateau, peaking in the steppe region of North Kazakhstan and in mountain region of Armenia (Balanovsky *et al*. 2015; Rootsi *et al*. 2012). The sister lineage G2-P287 engulfs the majority of the hg G lineages and is distributed over the whole range of hg G from Europe, the Near/Middle East and Central and South Asia with some founder sub-lineages among Jews and South Asians (Rootsi *et al*. 2012). The Tyrolean mummy Ötzi belongs also to the sub-clade of hg G in close affinity with the Sardinians (Keller *et al*. 2012).

Haplogroup **H-M2713** is autochthonous to India with a low frequency outside the sub-continent, and has relatively deep position within chrY phylogeny (Sengupta *et al*. 2006). Sequences from twelve South Asian individuals redefined this hg with the mutation M2713 (Figure 6), placing the previous diagnostic marker M69 more downstream (Poznik *et al*. 2016). Hg H and its sub-clades are largely confined to South Asia and comprise around quarter of the paternal lineages there (Sengupta *et al*. 2006). The variance distribution of H*-M69 and H2-Apt is higher in north and southeastern India, respectively, whereas the distribution peak of the variance of H1-M53 is in western India (Sengupta *et al*. 2006). The haplogroup is found also in Afghanistan among Pashtuns (3–8%), Tajiks (4–5.5%), and Hazara (2–12%) (Di Cristofaro *et al*. 2013; Haber *et al*. 2012). In Pakistan the reported frequencies are 7–9% (Di Cristofaro *et al*. 2013; Sengupta *et al*. 2006). Hg H-M69 has extremely low frequency in East and South East Asia, the Middle East and Europe (Figure 7) (Sengupta *et al*. 2006; Sahoo *et al*. 2006; Trivedi *et al*. 2008; Underhill *et al*. 2000; Gayden *et al*. 2007; Cadenas *et al*. 2008; Battaglia *et al*. 2009; Zhong *et al*. 2011; Grugni *et al*. 2012; Di Cristofaro *et al*. 2013; He *et al*. 2012). Outside of India the only notable presence of hg H is among the European Roma, who carry the lineage H-M82 with frequencies in different sub-populations ranging from 10–60% (Gresham *et al*. 2001; Gusmão *et al*. 2008; Klarić *et al*. 2009; Martínez-Cruz *et al*. 2016; Pamjav *et al*. 2011; Pericić *et al*. 2005; Regueiro *et al*. 2011). The widespread presence and low diversity of this South Asian specific lineage among groups of European Roma has been interpreted to vouch for their genetic origin from India (Gresham *et al*. 2001; Martínez-Cruz *et al*. 2016).

Next split within the global paternal phylogenetic tree is between the mega-haplogroup K-M9 and IJ-M429 cluster that branches further into two major haplogroups I-M170 and J-M304 (Figure 6) (Karafet *et al*. 2008; Poznik *et al*. 2016).

Haplogroup **I-M170** is a major European chrY lineage with the overall frequency of ~18% and is almost absent elsewhere (Rootsi *et al*. 2004; Semino *et al*. 2000; Underhill *et al*. 2007). This hg shows dual frequency peaks within Europe, it encompasses more than a third of paternal lineages in the Scandi-

navia (25–35%) and Balkans (17–40%), but is present at notable frequencies also in the eastern European Slavic-speaking populations (Figure 7) (Rootsi *et al*. 2004). Hg I-M170 has three main lineages. Two major sub-clades, I1-M253 and I2-PF3835 encompass majority of the variation that is proposed to reflect the continuity since Palaeolithic hunter-gatherers (Rootsi *et al*. 2004). I1-M253 has reached 25–35% among contemporary Scandinavian men (Rootsi *et al*. 2004). It is currently one of the three major paternal lineages (others being R1b-M269 and R1a-M198) in Europe taking up 14% of all paternal lineages, after having gone through a recent expansion about 5 KYA (Batini *et al*. 2015). Hg I was widespread among European hunter-gatherers (Fu *et al*. 2016; Günther *et al*. 2018; Lazaridis *et al*. 2014; Skoglund *et al*. 2014), the sub-lineage I3-L596 was shared between hunter-gatherers in Scandinavia (Mathieson *et al*. 2015) and Anatolian farmers (Lazaridis *et al*. 2016).

Haplogroup **J-M304** is present in the Near East, North Africa, Europe and India with the proposed origin in West Asia (Al-Zahery *et al*. 2011; Cinnioglu *et al*. 2004). In Crete, Turkey and Middle East (Azerbaijan, Oman, Syria, United Arab Emirates, Armenia) it reaches frequencies above 30% (Herrera *et al*. 2012; Luis *et al*. 2004; Nasidze *et al*. 2009), whereas in Egypt, Greece, Bulgaria and southern Italy the hg is present at 20–30% of the male lineage with a sharp drop towards northern and Central Europe (Di Giacomo *et al*. 2004). Hg J-M304 splits into two major lineages J1-M267 and J2-M172 that have an un-like distribution with J1-M267 restricted more to the Middle East and Arabian Peninsula and J2-M172 having a wider distribution reaching from the Balkans to eastern India, while populations in the Middle East have a roughly equal share of both lineages (Cinnioglu *et al*. 2004; Singh *et al*. 2016).

J1-M267 is proposed to have two distinct migratory episodes, the latest of which is linked to the spread of the Arab people (Chiaroni *et al*. 2010; Semino *et al*. 2004). J1-M267 is also present in many populations in the Caucasus region, reaching very high frequencies in some of them (Balanovsky *et al*. 2011; Yunusbayev *et al*. 2012). J1-P58, a prominent sub-clade of hg J1, is common in the Near East with the proposed origins in the Zagros/Taurus mountain region (Chiaroni *et al*. 2010). It has the highest frequency (74.1%) among the Marsh Arabs in Iraq and is the main paternal lineage in southwestern Mesopotamia (Al-Zahery *et al*. 2011). The most prevalent (46.1%) paternal lineage among the members of the Jewish priesthood is a haplotype nested within J1-P58 (Hammer *et al*. 2009).

J2-M172 is the more numerous sub-clade of hg J that differentiates into several subgroups with main division in to **J2a-M410** and **J2b-M12**. Based on the extant phylogeography it has been proposed that J2a-M410 and J2b-M12 are associated with demic diffusion of Neolithic farmers in North Africa and Eurasia from Mesopotamia (Iraq and Syria today) (Al-Zahery *et al*. 2011; Di Giacomo *et al*. 2004; Semino *et al*. 2004). **J2b-M12** has the highest frequency among Albanians (14.3%), it is the major male lineage in the Middle East and Mediterranean region with decreasing frequency towards Europe. Still, at low frequency it is widely spread from the Balkans to India and the Himalayas

(Semino *et al*. 2004). Within Indian hg J pool J1-M267 Y chromosomes are very rare, whilst most belong to J2-M172 (Singh *et al*. 2016). The two sister clades J2a-M410 and J2b-M102 make up a considerable part of the chrY gene pool of the Indian subcontinent. In total, the hg J frequency has been estimated to be around 5.1% (Sahoo *et al*. 2006; Sengupta *et al*. 2006; Trivedi *et al*. 2008), but frequencies may reach much higher values in some tribal populations (e.g. up to 77% in Asur), likely due to a founder-effect (Singh *et al*. 2016).

The megahaplogroup K-M9

Marker M9 is defining the megahaplogroup K that has sub-lineages on every continent (Figure 6, Figure 7). Within this megahaplogroup the initial bifurcation is between hg LT (known also as K1) and K2 or KxLT (K without LT) (Mendez *et al*. 2011; Poznik *et al*. 2016).

Haplogroups **L-M11** and **T-M184** have a distinctive geographic range where hg **L-M11** is spread within South Asia, in the Middle East, Caucasus and Anatolia and has a low presence among some Mediterranean European populations (Chiaroni *et al*. 2009; Poznik *et al*. 2016; Semino *et al*. 2000; Sengupta *et al*. 2006; Yunusbayev *et al*. 2012). The overall frequency of L-M11 in India is 5.6% whereas elsewhere its highest frequency is 4.3% in Central Asia (Trivedi *et al*. 2008; Wells *et al*. 2001). Hg **T-M184** is distributed with variable frequencies in west Asia, Africa, and Europe and has presumably Near Eastern origin. It is found among Near Eastern non-Jewish populations and is present in several Jewish populations (Batini *et al*. 2015; Mendez *et al*. 2011).

Recent sequencing efforts (Poznik *et al*. 2016) revealed a deeper substructure within hg **K2/K(xLT)-M526** revealing that together with some K-lineages prevalent in Indonesia and India (K2c,d,e), there is a subgroup K2a-M2308. The ancient sample Ust-Ishim (Fu *et al*. 2014) belongs to K2a*-M2308, upstream from NO-M214, a clade that contains two major lineages – hgs N-M231 and O-M175. The sister-clade to K2a-M2308 is K2b-P331 that encompasses hgs M,S, and the next megahaplogroup P-P295 (Figure 6).

Haplogroup **NO-M214** has two major sub-haplogroups with distinct and mostly non-overlapping geographic distribution – O-M175 and N-M231 (Rootsi *et al*. 2007), both with a likely region of origin in South East Asia (Hammer *et al*. 2006; Kayser *et al*. 2003; Rootsi *et al*. 2007), splitting from each other ~45 KYA (Poznik *et al*. 2016).

Hg **O-M175** is present at high frequencies in South East Asia, but also in Oceania, not reaching further west. It is by far the most prevailing hg in East Asia, comprising about 75% of Chinese and more than half of Japanese male population (Kayser *et al*. 2008; Shi *et al*. 2005; Wang and Li 2013; Yan *et al*. 2011). Hg **N-M231** ranges from the Beringian and Amur region in the Russian Far East across Siberia and northern China to eastern Europe (Zerjal *et al.*, 1997; Karafet *et al.*, 2002; Tambets *et al.*, 2004; Rootsi *et al.*, 2007). A recent whole chrY capture sequencing coupled with a comprehensive phylogeographic study revealed that the wide continuous distribution, which also cuts across linguistic and cultural borders, is divided into distinct regional sub-groups. In

particular, the clade N3a'6 reaches from Siberia to the eastern part of Fennoscandia and the eastern Baltic region with apparently nearly simultaneous and rapid regional diversification that took place within the last 5,000 years (Ilumäe *et al*. 2016).

Within hg K2, the variant M45 defines the **megahaplogroup P-M45** that besides other rarer lineages contains the hgs Q-M242 and R-M207. The other branch from K2, with diagnostic marker P331, leads to the two rare and mainly Melanesian lineages S and M. **Hg M-M104** is a haplogroup of suggested Melanesian origin (Kayser *et al*. 2001, 2003, 2006), M-M4 has 75% of the West New Guinea chrY lineages (Kayser *et al*. 2003). Hg S-M230 is a major hg in the Papua New Guinea, but also is found in Australia, Indonesian and Melanesian islands (Bergström *et al*. 2016; Kayser 2010; Kayser *et al*. 2006).

Haplogroup **Q-M242** is spread throughout Asia and has higher frequency in Central Siberia and the Americas. The first split within the haplogroup is between Q1'2-L472 and Q3-L257 that most likely occurred in Western or Central Asia (Balanovsky *et al*. 2017). Haplogroup Q2-F1096 is spread in Central and South East Asia, Siberia and its sub-lineage Q2b-B143 among Paleo-Eskimos (Balanovsky *et al*. 2017; Battaglia *et al*. 2013; Jota *et al*. 2016; Rasmussen *et al*. 2010). The other main branch, Q1-M364, is present in South Asia and Europe, Central Asia, Siberia and its subclades Q1a-M3 and Q1b-M971 encompass the majority of paternal lineages of the Americas (Dulik *et al.*, 2012; Battaglia *et al.*, 2013; Jota *et al.*, 2016). Haplogroup Q3-L257 is a more infrequent lineage present within western Eurasia and South Asia (Al-Zahery *et al*. 2011; Balanovsky *et al*. 2017; Grugni *et al*. 2012; Lacau *et al*. 2012; Lippold *et al*. 2014).

Haplogroup **R-M207** divides into two main lineages, R1-M173 and R2-M479 (Chiaroni *et al*. 2009; Karafet *et al*. 2008; Myres *et al*. 2011). The subclades of **R1-M173** are distributed widely over Eurasia from South Asia to western Europe. More than 50% of all European men carry Y chromosomes belonging to hg R1-M173 (Myres *et al*. 2011). Within Europe there is a sharp border between its two subgroups – hg R1a-M420 is mainly frequent in the eastern Europe, whereas R1b-M343 is most abundant in the western Europe (Batini *et al*. 2015; Kayser *et al*. 2004, 2005; Myres *et al*. 2011; Rosser *et al*. 2000; Semino *et al*. 2000; Underhill *et al*. 2015). **R1a-M420** carriers are widespread, reaching from Scandinavia to South Asia, but with clear separation between the Asian and European variants (Underhill *et al*. 2010, 2015). There are also branches characteristic to certain populations – like R1a-M582 among the Jewish groups (Rootsi *et al*. 2013) and R1a-Z284 in Scandinavia (Underhill *et al*. 2015). Also, European Roma carry R1a paternal lineages, notably Hungarian and Croatian Roma have the South Asian type (Chennakrishnaiah *et al*. 2013; Pamjav *et al*. 2011; Underhill *et al*. 2015). The sister clade **R1b-M343** is a hg of probable western Asian origin, with its subclade R1b-M269 carried to Europe. Today, more than 70% of the western European men bear sub-lineages of this hg (Batini *et al*. 2015; Poznik *et al*. 2016). There is a southeast to northwest frequency cline of R1b sub-lineages with higher frequencies in the

western Europe (Busby *et al*. 2012; Myres *et al*. 2011). A sub-lineage R1b-V88 is mostly found in Africa, although rare in the continent, it reaches higher frequencies in the central-western parts of it (Cruciani *et al*. 2010).

Haplogroup **R2-M479** has a much more restricted spread than R1-M173. It is one of the main autochthonous paternal lineages within India that are spread among different socio-linguistic groups and is likely present in the subcontinent already from the pre-Neolithic times (Sahoo *et al*. 2006; Sengupta *et al*. 2006; Trivedi *et al*. 2008). It has an almost uniform distribution all over the sub-continent, but a notable cline is detected among different castes, with a decline in frequency toward upper castes (Trivedi *et al*. 2008). Hg R2-M124 frequency in India is 8–13.5%, reaches 7.4% in Pakistan, and decreases towards Central Asia (3.8%) (Cordaux 2004; Di Cristofaro *et al*. 2013; Kivisild *et al*. 2003; Sengupta *et al*. 2006; Trivedi *et al*. 2008; Wells *et al*. 2001).

In sum, thus far, the extant global variation of chrY lineages is described with fine detail, coupled with temporal scale. The same applies for mtDNA, not described in this thesis. The finely mapped geographic distributions of paternal and maternal lineages show distinct global and local distribution patterns on population level. Currently, similar databases are growing for individuals passed away hundreds and thousands of years ago – the study of ancient DNA is thriving.

2.6 Perspectives from ancient DNA

The possibility to sequence DNA directly from ancient human remains, the ancient DNA (aDNA), allows direct assessment of the genetic variation from a specific location and time in the past. The specific challenges of aDNA study are the low amount of surviving material available for DNA extraction, often very low endogenous DNA content, DNA degradation and specific changes to the DNA bases (Orlando *et al*. 2015). There are significant advances in the possibilities to sequence stretches of DNA that are shorter than PCR fragments, methods for endogamous ancient DNA detection, enrichment of specific SNVs with the targeted capture methods, and data analyses. This has resulted in the ever accumulating knowledge from aDNA that reshapes and widens our understanding of human demographic history (Haak *et al*. 2015; Lazaridis *et al*. 2016; Mathieson *et al*. 2015; Orlando *et al*. 2015; Schlebusch *et al*. 2017). For example, emerging knowledge about Sub-Saharan African aDNA has supported the claims of the population structure within Africa and also pushed back the time estimate of the divergence between Khoisan populations and the rest of modern human populations to have occurred ~350–260 KYA (Schlebusch *et al*. 2017). Another study found corroborating African aDNA evidence for the Levantine farmers' back migration to Africa (Skoglund *et al*. 2017), an ancestry component until then observed among the contemporary individuals (Pagani *et al*. 2012; Pickrell *et al*. 2014; Tishkoff *et al*. 2009).

Prior to aDNA studies, the prevailing broad view to European demographic history, based on the reconstructions from contemporary uniparental lineages, was that Europe was populated by two main migration waves: the initial arrival of hunter-gatherers and their subsequent replacement by the Neolithic farmers arriving from the Near East. These major events were topped with additional post-glacial recolonizations and inter-continental movements during the Bronze Age that also could have obliterated the traces from earlier events (Soares *et al*. 2010). The recent aDNA results have confirmed some, but also added more complex details to the views of European population history, summarized into three major migration waves reaching Europe during the Upper Palaeolithic, the Neolithic and the Early Bronze Age (Allentoft *et al*. 2015; Haak *et al*. 2015; Lazaridis *et al*. 2014). The aDNA analyses on a more local scale have shown that the process of Neolithic transition of replacement and admixture between populations of diverse cultural background in Europe was regionally structured. Similarly to classical studies (Ammerman and Cavalli-Sforza 1973), the aDNA research has shown that Neolithic revolution in southern and Central Europe started with the arrival of early farmers from the Near East (Haak *et al*. 2015) who reached as far as Scandinavia (Skoglund *et al*. 2012; 2014). On the other hand, the intensive cultivation and animal husbandry was brought to the Eastern Baltic region of Europe by the people of mainly steppe ancestry (Mittnik *et al*. 2018; Saag *et al*. 2017).

Similarly, aDNA results from the human remains belonging to the Lapita Culture in South Pacific revealed that the initial colonisation of the remote Oceania was not by a mixture of populations from New Guinea as previously thought. The Papuan ancestry seen today in the remote Oceania represents later human movements (Skoglund *et al*. 2016).

Studies on ancient Y chromosomes

The ancient chrY (aY) sequence and genotype data is accumulating. This gives a possibility to directly test the hypotheses until now posed on the extant genetic variation data, often linking the spread of specific chrY hgs together with certain archaeological cultures. The placement of an ancient individual's chrY into the existing phylogeny reveals their relations with other ancient and extant specimens. This allows for the investigation of geographic and temporal changes in the composition of the lineages as well as the population sizes through time, indicative of migrations and other significant demographic events. An aY with substantial sequencing coverage also provides a way for calculation of the mutation rate through counting the mutations that have accumulated in modern lineages after the split from the ancient lineage (Figure 3 a.3.) (Fu *et al*. 2014; Rasmussen *et al*. 2014). It is still challenging and costly to obtain high coverage of chrY from the whole genome sequencing of the ancient specimen because of the complicated repetitive structure of chrY, its homology to the chrX, and the haploidy. These challenges are to some extent compensated by large number of redundant SNVs mapping to the same branch of the phylogeny that provide information to determine hgs also at moderate

coverage of the whole genome sequences from ancient specimen (Kivisild 2017; Van Oven *et al*. 2014; Rasmussen *et al*. 2014). It is also helpful that the hybridization-based methods targeting genome-wide SNVs now include over 30K of chrY markers for high resolution typing of aY variants (Lazaridis *et al*. 2016; Mathieson *et al*. 2015).

Even though aY sequences are emerging also from Sub-Saharan Africa (Schlebusch *et al*. 2017; Skoglund *et al*. 2017), the majority of the aY results have accumulated from the geographic regions in higher latitudes (northern/ temperate Eurasia and America) that favour better DNA preservation over long periods of time (Hofreiter *et al*. 2015). As reviewed recently in (Kivisild 2017), several conclusions based on extant chrY have been corroborated by aY, but others have not found support. All the oldest aY lineages (Fu *et al*. 2014, 2015, 2016; Seguin-Orlando *et al*. 2014) from Eurasia belong to the three main basal lineages, C, D and F, documented as a result of the OOA movement, and coalesce ~50–20 KYA, just as was predicted from the distribution of the extant lineages. Finding of hg C, nowadays extremely rare in Europe, among the Iberian hunter-gatherers (Olalde *et al*. 2014), and Anatolian and central European Neolithic farmers (Mathieson *et al*. 2015) reveals that the geographic distributions of paternal lineages have gone through major changes through time. Extant distribution of hg G suggested that it was brought to Europe by the Anatolian farmers (Semino *et al*. 2000) and aY results corroborate that (Broushaki *et al*. 2016; Hofmanová *et al*. 2016; Lazaridis *et al*. 2016; Mathieson *et al*. 2015). Hg I, in contrast, was proposed to be the sign of local population continuity from Palaeolithic hunter-gatherers (Rootsi *et al*. 2004). Indeed, it is found among the remains of Palaeolithic hunter-gatherers (Fu *et al*. 2016; Günther *et al*. 2018; Lazaridis *et al*. 2014), in some cases showing some regional continuity, but it is also present among the Anatolian farmers (Lazaridis *et al*. 2016). So, the aY draws a more complex picture than previously thought. Hg J, the sister lineage to hg I, was thought to have arrived to Europe with the early Neolithic farmers (Rosser *et al*. 2000; Semino *et al*. 2000). It is not detected from European Neolithic remains thus far, instead it is found among hunter-gatherers in distant regions in the Caucasus and Karelia in eastern Fennoscandia, and among early farmers in Iran and Anatolia (Jones *et al*. 2015; Lazaridis *et al*. 2016; Mathieson *et al*. 2015). It follows that hg J reached Europe only during the Early Bronze Age through the Yamnaya/Corded Ware migrations (Allentoft *et al*. 2015; Haak *et al*. 2015). Ancient chrY results show that the nowadays most frequent lineage in Europe, R1b, contrary to the early view of being a Mesolithic remnant (Rosser *et al*. 2000; Semino *et al*. 2000), became common in Europe only during the Late Neolithic and the Early Bronze Age (Allentoft *et al*. 2015; Haak *et al*. 2015; Mathieson *et al*. 2015). A large proportion of the Bronze Age aY lineages from Central Europe, northern Caucasus and Steppe Belt belong to either R1a or R1b lineages, but show regionally characteristic clustering within these major hgs (Allentoft *et al*. 2015; Cassidy *et al*. 2016; Haak *et al*. 2015; Mathieson *et al*. 2015; Schiffels *et al*. 2016). There might be rare lineages that have persisted longer, but the majority

of European lineages stem from only a few Late Neolithic early Bronze Age R1a and R1b founders, a fact also seen from the sequencing data of extant European Y chromosomes showing the coalescence of the main branches to be shallow around 5–7 KYA (Batini *et al*. 2015; Hallast *et al*. 2015; Poznik *et al*. 2016).

The majority of extant American lineages belong to two major hgs Q1a-M3 and Q1b-M971 (Jota *et al*. 2016; Zegura *et al*. 2004), and minor lineages to hg C3-M217. The unequal distribution of these lineages within the continents has given support for the dual origin model of Native Americans in the sense of separate migration waves reaching the Americas (Lell *et al*. 2002; Roewer *et al*. 2013). On the other hand, the observed similar levels of STR diversity among the lineages argues for a model of a single migration wave 10–17 KYA (Zegura *et al*. 2004). Ancient Y chromosomes have shown the presence of Q1a-M3 and Q1b-M971 in the Americas at least from 10.3 KYA (Kemp *et al*. 2007; Rasmussen *et al*. 2014, 2015). Ancient DNA from the Saqqaq man from Greenland (Rasmussen *et al*. 2010) gave direct evidence for a separate Paleo-Eskimo dispersal into the Arctic of North America, the aY lineage he carried is Q2b-B143, a hg that might have extant descendants (Matthew C Dulik *et al*. 2012; Olofsson *et al*. 2015), despite the complete population replacement by the Neo-Eskimos.

African aDNA is still scarce, but the Y chromosomes of four ~2 KY old Stone Age hunter-gatherers from South Africa belong to hg A1b1b2a (aka A3b1-M51) (Schlebusch *et al*. 2017; Skoglund *et al*. 2017). This hg is common among present-day Khoe-San in Southern Africa (Barbieri *et al*. 2016), thus providing evidence for population continuity. The aY of 6–8 KY old hunter-gatherers of Malawi have been assigned to hg BT, whereas a 400 Y old Kenyan has hg E1b1b1b2 (Skoglund *et al*. 2017), and also a 4.5 KY old Ethiopian individual carried hg E1b1 (Gallego Llorente *et al*. 2015), representing the most widely spread hg in Africa today.

2.7 Examples of founder populations – European Roma and Ashkenazi Levites

South Asia and its European offshoots, the Roma people

South Asia is comprised of India, Nepal, Pakistan and Bangladesh and is one of the most densely populated regions in the world. The 2011 census of India counted 1.2 billion people while South Asia together is home to 1.75 billion people. South Asian people are culturally, morphologically, linguistically and genetically highly diverse and interact in intricate and complex social structures. About 8% of the people in India are classified as ‘tribal’ and their major modes of subsistence are hunting and gathering, foraging and seasonal agriculture of various kinds. The majority of Indian populations are the ‘non-tribals’ who have been classified socially into castes under the Hindu Varna system or

by their religious affiliations (i.e., Christian, Muslim or other). Caste system has a very elaborate social structure that evolved from a system of the traditional occupations (Mukerjee 1937). Since the caste system has strongly stratified Indian society and created severe inequality, there are now affirmative actions for socially backward castes and tribes who are officially designated as Scheduled Castes and Scheduled Tribes. The human population genetic studies of Indian populations generally group individuals on bases of their self-reported group affiliations. These are usually based on either the Varna system (the rank of castes or tribe-caste dichotomy), or other socio-cultural factors like language, means of subsistence and traditional customs. Besides, groups are formed purely based on geographic or linguistic affiliations. Languages spoken in India belong to four linguistic families – Indo-European, Dravidian, Tibeto-Burman and Austro-Asiatic, all with their distinct distributions. The most widely spoken languages belong to the Indo-Aryan language group (from Indo-European linguistic family).

European Roma, today around 11 million people scattered in small groups throughout Europe, are the only speakers of an Indo-Aryan language in this region (Cahn and Guild 2008). The Romani language belongs to the central Indo-Aryan subgroup, just as Hindi language. It also has traces from Iranian languages and Armenian, with profound impact from Greek, topped by the influence from the European host population languages and local dialect formations (Matras 1999, 2005; Turner 1926). The substantial scholarly interest in the Roma, the linguistic and cultural outliers in Europe, has resulted in a substantial amount of literature of linguistic, anthropological and ethnological aspects of this peripatetic population, giving the understanding that European Roma originate from India. For the lack of written and oral histories of their own, their origin from India and subsequent migrations were revealed first by linguistic and anthropologic evidence (Fraser 1995; Matras 1999; Rüdiger 1782) and later corroborated by genetic studies (Bánfai *et al*. 2018; Gresham *et al*. 2001; Kalaydjieva *et al*. 2005; Melegh *et al*. 2017; Mendizabal *et al*. 2012; Moorjani 2013). The suggested dates for the exodus from India are around 10^{th}–11^{th} century and arrival to Europe soon after. Local chronicles throughout Europe have recorded the subsequent spread by14th–15th century (Fraser 1995). Roma today have substructured into series of endogamous populations spanning country borders, their exclusionary culture and hostility from the surrounding populations has often kept them stigmatized and in the periphery of their host societies (Cahn and Guild 2008; Hancock 2002).

Uniparental lineages of European Roma can be largely classified as originating from South Asia or West Eurasia. Reported South Asian contribution to the maternal gene pool is restricted to only a very few lineages (M5a1b, M35b, M18). The bulk of those belong to mtDNA hg M5a1b that is detected among all studied Roma populations at 6–30%. Other two South Asian mtDNA lineages are present more sporadically and are much more infrequent (Gresham *et al*. 2001; Gusmão *et al*. 2008; Malyarchuk *et al*. 2006; Mendizabal *et al*. 2011; Salihović *et al*. 2011). The main South Asian contribution to the paternal

lineages is chrY hg H1a-M82, reported thus far in all studied Roma populations. The frequency of this hg varies between 10–48%, with clinal frequency change from the Balkan Roma towards Iberian Roma (Gresham *et al.*, 2001; Gusmão *et al.*, 2008; Klarić *et al.*, 2009; Zalán *et al.*, 2011; Regueiro *et al.*, 2012). Another paternal lineage with current South Asian distribution R1a-M780 is thus far identified among Hungarian and Croatian Roma (Chennakrishnaiah *et al*. 2013; Pamjav *et al*. 2012; Underhill *et al*. 2015). Besides the South Asian founder lineages also a number of Western Eurasian specific paternal and maternal founder lineages have been identified (Martínez-Cruz *et al*. 2016).

The genome-wide SNV analyses uncovered a general profile of Western Eurasian and South Asian ancestry components among the genomes of European Roma. The West Eurasian genetic component has been demonstrated by high average sharing of identical-by-descent (IBD) genomic segments with European populations (Melegh *et al*. 2017; Mendizabal *et al*. 2012; Moorjani *et al*. 2013) and with populations from the Caucasus region (Bánfai *et al*. 2018). The high heterogeneity of West Eurasian uniparental lineages among Roma advocates a continuous admixture, in sharp contrast to the few South Asian founder lineages with very limited diversity spread at varying levels among all the Roma populations (Gresham *et al*. 2001; Kalaydjieva *et al*. 2005; Martínez-Cruz *et al*. 2016; Mendizabal *et al*. 2011, 2012; Moorjani *et al*. 2013; Morar *et al*. 2004). The population genetic studies of the classical, uniparental and autosomal genetic variation of European Roma have concluded that Roma stem from a single founder population that has subsequently gone through strong drift and admixture events and a series of local founder events. Small populations/groups constitute the European Roma population as a whole. Medical genetic studies have identified a number of group-specific private founder mutations as causes of rare conditions among Roma individuals (Barešić and Peričić Salihović 2014; Fiore *et al*. 2011; Kalaydjieva *et al*. 1999; Kalaydjieva, Gresham, *et al*. 2001). The genetic structure within the contemporary European Roma as a single group can be best explained by their main migratory movements within Europe (Gresham *et al*. 2001; Gusmão *et al*. 2008; Kalaydjieva, Calafell, *et al*. 2001; Kalaydjieva, Gresham, *et al*. 2001; Klarić *et al*. 2009; Malyarchuk *et al*. 2006, 2008; Martínez-Cruz *et al*. 2016; Mendizabal *et al*. 2012; Moorjani *et al*. 2013; Salihović *et al*. 2011).

One of the long-lasting gripping questions has been the possible region of origin within South Asia for the ancestors of European Roma. Northwestern India has been overwhelmingly pointed to as the main source for the South Asian genetic component among the European Roma (Kalaydjieva, Gresham, *et al*. 2001; Melegh *et al*. 2017; Mendizabal *et al*. 2011; Moorjani *et al*. 2013), in addition also East India has been indicated for mtDNA (Mendizabal *et al*. 2011) and South India for chrY (Regueiro *et al*. 2011) as potential source regions. Analyses of autosomal markers affiliated Punjabis, Meghaval and Kashmiri Pandits as the closest relative contemporary South Asian populations to the European Roma (Mendizabal *et al*. 2012; Moorjani *et al*. 2013). On the other hand, Punjabis and Kashmiri Pandits were also the populations with the highest

West Eurasian admixture among Indian populations which could be a confounding factor (Moorjani *et al*. 2013). A recent reanalysis concludes that the autosomal South Asian component among the European Roma shares the highest amount of IBD-segments with the populations from northwestern India, but adds also populations of Pakistan as a main genetic contributor (Melegh *et al*. 2017). A major drawback in the comparability of different studies has been that the grouping of Indian populations into geographical categories is not always consistent across studies and, especially in the earlier research on uniparental lineages, sampling of the Indian populations was underrepresenting the genetic variation within India, leading to potential biases.

A brief overview of history, demography and genetic studies of the Jewish people

Notes on the history of the Jewish people

Jews are an ethno-religious Semitic-speaking group with current census size of about 14.3 million. Historic records place their origins in the Levant at around second millennium before the Common Era (BCE) (DellaPergola 2015). Due to numerous exiles, persecutions and displacements in different times of their history, they have been dispersed into various groups, residing in many parts of the world, a phenomenon known as the Diaspora (DellaPergola 2015; Ostrer 2001). Despite this, Jews have kept strong cultural, traditional and genetic continuity, the main characteristics of different communities are the migration history, place of long term residence and some varieties in the customs. Turbulent times of Babylonian and Persian Empires (4^{th}–6^{th} centuries BCE) gave rise to the Middle Eastern Jewish communities, the *Mizrahi* (East in Hebrew) *Jews*. Classical Antiquity (8^{th} century BCE to 6^{th} century AD) saw the formation of the Balkan, Italian, North African and Syrian Jewish communities that later mixed with the Sephardi Jews. *Sephardi Jews* initially lived in the Iberian Peninsula, where they reached in larger numbers during the beginning of the Common Era after Romans had destroyed Jerusalem at 70 CE and upheavals that followed (Bowers 1975; Noy 2013). Many Jews made it to Iberia also later as part of the Muslim expansion (starting in 7^{th} century CE). After their expulsion from Iberia at the end of 15^{th} century they migrated predominately to North Africa and to the western Asian and European parts of the Ottoman Empire, and at lesser extent to West Europe, and mixed with other Jewish groups there. One of the most widely known scenarios about the immediate ancestors of the *Ashkenazi Jews* suggests that groups of Jews migrated from northern Italy and settled in the Rhine Valley around 7^{th}–9^{th} centuries. There formed a distinct population, later known as *Ashkenazi Jews*. Subsequently, between 11^{th} and 15^{th} century, many of them migrated further to Central and East Europe – Poland, Bohemia, Lithuania and further east. In addition, there are several other distinct Jewish communities scattered in many parts of the world. The strong cultural unity of distant Jewish communities have facilitated a considerable amount of movement of people between them throughout time with the main distinct

religious characteristics of the groups still persisting. World Jewry today can largely be divided into Ashkenazi and non-Ashkenazi Jews, whereas roughly two thirds of all Jews are the descendants of the Ashkenazi group (DellaPergola 2015).

Throughout the Jewry World, the paternal lineages can be divided into three castes – Israelite, Cohen (plural Cohanim) and Levite. The latter two, each representing approximately 4% of the Jewish males, were historically priestly castes. Priests who by tradition are said to be the descendants of the Biblical forefathers Levi and his great-grandson Aaron, had distinctive rights and rules to obey and carried an important role in the Jewish culture. The given name of the forefather Levi eventually transformed into an honorific name "*Halevi*" ('the Levite', in Hebrew) for the whole tribe of priests. In turn, the tribe of high priests, descending straight from Levi's great-grandson Aaron, were given the honorary name that represented their occupation – Cohen or "*Kohanim*" ('priest', 'to serve' in Hebrew). The present-day Jewish religious laws determine Jewishness as a maternally determined trait – transmitted by a Jewish mother to her children or alternatively, and more rarely, can be obtained through official rabbinical conversion. The same law demands that the Levite and Cohen statuses have a direct patrilineal descent, eventually leading to the respective Biblical forefathers (Anon 1972). Thus the transmission of the cultural traits of Cohen and Levite statuses resemble the inheritance mode chrY and therefore present exemplary cases for the study of coinheritance of biological and cultural traits.

The Cohen and Levite statuses have transformed to be the surnames of their carriers, so the surnames are additional markers for the cultural trait. Through time these names have evolved into many varieties, Levi has become e.g. Levin, Lewicki, Loewi; whereas Cohen has changed into e.g. Kogan, Kahn, Kohanuy. When taking the names as signifying the respective status, one has to consider that the carriers of Levite or Cohen status have also taken other surnames, for example toponymic surnames – names related to the location of residency of the respective family. This is exemplified by the Horowitz Rabbinical Levite dynasty, whose members carry the name related to the Bohemian town of Hořovice where the founder of the dynasty settled after emigrating from Catalonia in 15th century (Horowitz 1928). This Levitical dynasty is also a well-known Rabbinical court maintaining a detailed documented genealogy. The many meticulously preserved historic records and deep interest in the Horowitz family history fuels the contemporary thriving research of family lines by enthusiastic genealogists, today also supported by genetic tests (Henn *et al*. 2012), a way which, when properly interpreted (Falk 2015; Hammer *et al*. 2009; Tofanelli *et al*. 2016), can lead to fruitful collaboration between academic and citizen scientists (Balanovsky *et al*. 2017) besides discoveries in the personal family history.

Genetic studies of the Jewish populations

The genetic affinities of several Jewish populations to each other and to their neighbouring populations have been repeatedly studied (Carmelli and Cavalli-Sforza 1979; Efron 2013; Karlin *et al*. 1979; Salaman 1911). The intricate population histories, bottlenecks, founder events and endogamy of different Jewish groups manifest also through distinctive prevalence patterns of several Mendelian diseases caused by population specific founder mutations. The extensively studied Jewish medical genetics and its fruitful translation into public health initiatives (Baskovich *et al*. 2016) are beyond the scope of this overview, for a brief review see (Ostrer and Skorecki 2013).

The main questions studied with respect to Jewish population genetics have primarily addressed the extent of the commonly shared gene pool among different Jewish groups and the admixture proportions with surrounding populations. The early studies on blood groups and serum markers showed that distinct Diaspora Jewish groups have common Middle Eastern origin, showing closer affinity to each other than to their surrounding populations, and have differential level of admixture with the surrounding populations (Carmelli and Cavalli-Sforza 1979; Karlin *et al*. 1979; Livshits *et al*. 1991). The following studies on the variation of Y chromosomes and mtDNA also showed founder effects of Middle Eastern and local origin (Behar *et al*. 2003, 2004; Hammer *et al*. 2009). These principle findings have been corroborated with more nuanced datasets and different analysis methods recently, by studies of genome-wide variation among different Jewish groups (Atzmon *et al*. 2010; Behar *et al*. 2010, 2013; Bray *et al*. 2010; Carmi *et al*. 2014; Kopelman *et al*. 2009; Xue *et al*. 2017).

Genetic studies of Ashkenazi Jews

The Ashkenazi Jews constitute the most numerous of the Jewish communities and their origins have been extensively studied for over a century (Fishberg 1911; Ostrer and Skorecki 2013). The genome-wide studies of contemporary Ashkenazi Jews have identified, similarly to other Jewish populations, the Middle Eastern ancestral sources, but with relatively higher European admixture and shown the close genetic relatedness between Ashkenazi Jews and other Jewish groups (Atzmon *et al*. 2010; Behar *et al*. 2010; Bray *et al*. 2010; Carmi *et al*. 2014). It has been shown that Ashkenazi Jewish population defies the simple models of genetic history, for the population has undergone founder events, expansions with intermittent bottlenecks and several admixture events (Bray *et al*. 2010; Carmi *et al*. 2014; Palamara *et al*. 2012; Risch *et al*. 2003; Xue *et al*. 2017). From the genome-wide data the amount of European admixture has been estimated to be around 35–60% (Atzmon *et al*. 2010; Bray *et al*. 2010; Xue *et al*. 2017).

Studies of uniparental marker systems have revealed specific founder lineages and introgressed lineages from surrounding populations. A strong founder effect is suggested for the Ashkenazi Jewish maternal lineages coalescing to only four major Near Eastern mtDNA founding lineages K1a1b1a, K1a9, K2a2a1 (32% in total) and N1b2 (9%) that are otherwise rare in Europe and the Near East

(Behar *et al*. 2006). The much disputed 'Khazar-hypothesis' of the Ashkenazi origin – a substantial contribution of people from the Turkic Khazar kingdom to the Ashkenazi population (Elhaik 2013), has not found support from further genetic studies (Behar *et al*. 2013; Flegontov *et al*. 2016). Interestingly, the presence of two paternal hg J lineages, J1a-P58 and J2-M12, that are frequent among contemporary Jews, were identified in two ancient Canaanite samples that date to about 3.7 KYA. This provides the approximate time-estimate for these lineage in the Middle Eastern regions of historic Jewish settlements (Haber *et al*. 2017).

Paternal lineages among Ashkenazi Jews

The paternal lineages of Ashkenazi Jews have a more diverse set of hgs, but with less allelic diversity of STRs within haplogroups, than their European non-Jewish neighbouring populations. This lends support to the non-European component of the Ashkenazi Jewish origin (Behar *et al*. 2004; Thomas *et al*. 2002). Of the 19 hgs present among Ashkenazi Jews, only seven (E-M35, G-M201, J1-M267, J2-M172, Q-M242, R1a1-M17, R1b-P25) are reported above frequencies of 5%, but these seven comprise ~85% of their paternal lineages (Balanovsky *et al*. 2017; Behar *et al*. 2003). Out of the hgs listed above, E-M35 and J-M267, are spread among both, the Ashkenazi Jews and contemporary Middle Eastern populations (16% and 19% respectively). At the same time these lineages are present at much lower frequencies among non-Jewish European populations (~1% for both), thus suggested to be among the major founding lineages among Ashkenazim. The considered minor founder lineages, such as G-M201 and Q-P36, are present among Ashkenazi at a lower level (~8% and 5%), but with a distribution similar to that of the major founding lineages. The main lineages presumably representing European introgression to the Ashkenazi Jewish community, based on their current distribution and known phylogeny at the time of the study, were I-P19, R1b-P25, and R1a1-M17. Though the latter two are also present in the Middle Eastern non-Jewish populations, so they could have been among the initial founding lineages (Behar *et al*. 2004; Hammer *et al*. 2000; Nebel *et al*. 2001). The low haplotypic STR diversity among Ashkenazi Jewish paternal lineages is in concordance with the decreased number of mtDNA diversity and presence of the high level of recessive disease alleles among the population, signalling the reduction of effective population size and subsequent drift in a population after bottlenecks and founder effects (Behar *et al*. 2003, 2004, 2006; Ostrer and Skorecki 2013).

The variation of paternal lineages among the Jewish priestly castes

The studies on the Y chromosomes of the members of both Jewish priestly castes have revealed that a large fraction of Y chromosomes of Cohanim constitute a tight cluster irrespective of their Ashkenazi or Sephardi affiliations. Within the dominant (46.1% of all Cohanim lineages) hg J-P58 lineage with likely Near Eastern origin, 64.6% Cohanim in this haplogroup cluster around

the extended Cohen Modal STR Haplotype (Hammer *et al*. 2009; Skorecki *et al*. 1997). The Levites who comprise 4% of all Jewish men show signs of multiple genetic origins by having higher haplogroup diversity among them. The Ashkenazi Levites, on the other hand, carry another dominant (frequency over 50%) paternal lineage.

At the time of the identification, this lineage was classified as R1a1-M17, with frequencies much higher among Ashkenazi Levites than in any other Jewish group (Skorecki *et al*. 1997; Thomas *et al*. 1998). An early focused study on the Ashkenazi Levites found that the STR profiles of Levites within the R1a1-M17 lineage cluster tightly, testifying for a recent common ancestor within this lineage as opposed to the multiple introgression events (Behar *et al*. 2003, 2004; Nebel *et al*. 2005). The rarity of R1a1-M17 hg among the other Jewish people and Middle Eastern non-Jewish people and the abundance of this lineage among the Europeans, in particular Eastern European Slavic-speaking populations, suggested that its presence among the Ashkenazi Levites is likely a result of introgression from the surrounding "host" populations. The coalescence age estimate of hg R1a1-M17 and its low STR diversity within Ashkenazi Levites coupled with their history of continuous residence among the Eastern Europeans led initially to the conclusion that there was a single founding event of European introgression to the Ashkenazi Levite population (Behar *et al*. 2003).

The lack of information on the internal substructure of hg R1a did not allow more detailed study of the potential source. With the advent of chrY sequencing, it was possible to more precisely determine the phylogeny R1 clade and that of the R1a-M17 lineages, by fully sequencing 19 Y chromosomes (Rootsi *et al*. 2013). Coupled with the population based survey of distinctive sub-branch, it was revealed that all the R1a-M17 chromosomes from Ashkenazi Levites, together with other Middle Eastern R1a-M17 samples, formed a distinct sub-clade within R1a diversity, separated from the mainly eastern European clades. The geographic distribution of the diagnostic mutation of this branch, M582, showed that all Ashkenazi Levite R1a Y chromosomes carry it. The same mutation has a frequency of 33.8% among non-Levite Ashkenazi R1a males and 5.9% in Near Eastern R1a Y chromosomes, but was not sampled among European host populations (Rootsi *et al*. 2013).

Hence, the fully sequenced Y chromosomes showed that the previously suggested hypotheses about eastern European introgression to the Ashkenazi Levites was premature. Since the study of (Rootsi *et al*. 2013) included only two resequenced Ashkenazi Jewish Y chromosomes, a more comprehensive dating and internal diversity of the sub-clade remained somewhat obscure.

3. AIMS OF THE STUDY

In this thesis I study the genetic diversity of human Y chromosome on a global and regional scale and analyse the male-specific aspects of demography including dynamics of effective population size, (male) population migrations and origins in a range of geographical regions. Technological advancement has recently made it possible to resequence rather than merely genotype chrY, contributing significantly to the potential of human Y chromosome as a tool in population genetic studies. Harnessing these fruits I focus on the following scientific problems.

First (REF I), to characterise the global chrY phylogeny and male-specific aspects of population size dynamics from high-coverage short read sequencing data:

- Determine the accessible regions of chrY with the short read sequencing approach.
- Calibrate the chrY mutation rate with two ancient chrY.
- Reconstruct the dated phylogenetic tree with the highest possible resolution and characterise the global and local expansions of paternal lineages.
- Analyse the dynamics of male and female effective population sizes through time.

Second (REF II), to elucidate the ancestral relationship between an ancient, 24,000-years-old individual MA-1 from southern Siberia and contemporary humans. For that:

- Characterize the uniparental lineages of the MA-1 in the context of contemporary lineages and contextualize with the whole genome data.

Third (REF III), to improve the phylogeographic resolution of chrY hg H1a1a-M82 based on the STR diversity and locate the potential ancestral source of European Roma paternal lineages within India.

Fourth (REF IV), to fully characterize the chrY lineage R1a-Y2619 highly prevalent among Ashkenazi Levites. For that:

- Date the coalescence of the clade
- Determine the potential geographical region of origin of the lineage
- Compare the genetic pattern of R1a-Y2619 lineage to other common lineages among Ashkenazi Jews to discern if the founder event is Levite specific.

4. MATERIALS AND METHODS

The origin of the human DNA samples analysed in the present study, along with the experimental and analytical methods used, are described in detail in the respective research articles and/or their supporting materials.

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

First study seeks to define the parts of chrY reachable with SRS technologies by sequencing 456 worldwide samples, and then, with this highest possible resolution at hand, to uncover further the intricate details of the global paternal phylogenetic tree and comparative dynamics of male and female effective population sizes (REF I). Next study examines the ancestral relationships between an ancient individual and contemporary human populations, by analysing the low coverage sequence of the genome, including the Y chromosome and mtDNA, of a 24,000-year-old individual from southern Siberian archaeological site Mal'ta (REF II). The following two case studies focus on specific chrY hgs that show signs of founder events in particular populations. Within the extensive South Asian sample set, we examine the STR diversity and phylogeography of chrY hg H1a1a-M82, a South Asian specific autochthonous hg that is also a founder paternal lineage among the European Roma population (REF III). The fourth study harnesses the potential of the capture and sequencing of chrY in characterizing in detail the origin and spread of the haplogroup R1a-Y2619, a chrY sub-lineage with a strong founder event among Ashkenazi Levites (REF IV).

5. RESULTS AND DISCUSSION

This section is a concentrate of the results from four scientific articles that form the original work for the thesis. Here the aim is a comprehensive but concise overview of these results, with the details in the original articles and in their supplementary materials.

5.1 Global variation of Y chromosome reveals a recent bottleneck in male effective population size (REF I)

Sequencing the human Y chromosome

In this study we aimed to uncover with SRS as much as possible of the chrY diversity, construct the underlying phylogeny and estimate the temporal dynamics of the corresponding male and female effective population sizes from a sample of 456 individuals. The sequence structure of chrY is complicated (Skaletsky *et al*. 2003), so we developed several filtering schemes to first define the regions approachable with SRS technology. The most efficient filters in removing low quality sites were the "re-mapping filter" which excludes all regions of chrY that evade correct mapping of short reads and the "normalized coverage filter" that excludes regions with high fluctuation of unique coverage in the sample (SI 2, SI Table S2 REF I). After all the filtering steps we had 8.8 Mb of chrY sequence per individual and a total of 35,700 SNVs (SI Table S8 REF I) to be used in the further phylogenetic analyses and calculation of coalescence ages. We calculated the mutation rate in the filtered regions to be 0.74×10^{-9} (95%CI = $0.63–0.95 \times 10^{-9}$) per bp per year (SI 3 REF I) using data from two independent ancient DNA samples (Rasmussen *et al*. 2010, 2014) and the approach described in (Rasmussen *et al*. 2014).

The global human Y chromosome phylogeny

We rooted the phylogeny (Figure 1; SI Figure S3 REFI) with two hg A00 (Mendez *et al*. 2013) sequences and found the coalescence of hg A00 and the rest of the human chrY lineages (A2'5) to be about 254 KYA (Supplemental Table S7, REF I). There are 236 markers separating the rest of the phylogeny from the strictly African lineages A and B (Supplemental Fig. S13, S14, S15 REF I). There is a more than 15 KY gap between the first split at 68–72 KYA of the all non-African lineages and their subsequent differentiation into major hgs now common in Eurasia, America and Oceania at 47–52 KYA (SI Table S7, SI Figure S9 REF I). Within these non-African lineages first hg C shows bifurcation into C3 and all other C lineages (SI Figure S20, S21, REF I). Among the common non-African hgs just single mutations define splits between hgs F, GT and HT (Supplemental Figure S13, REF I). The subsequent inner branches (IT, K, NR, MR, P) are also short (SI Fig. S9 REF I) and in

agreement with the fast diversification of basic Eurasian and Oceanian founder lineages at around 50 KYA when these regions were colonized by anatomically modern humans (Bowler *et al*. 2003; Higham *et al*. 2014; Lippold *et al*. 2014; Mellars *et al*. 2013; Poznik *et al*. 2016). At the same time these results from a single genetic locus with extremely low N_e (Lippold *et al*. 2014) cannot refute the Middle Pleistocene dispersals from Africa along the southern route (Armitage *et al*. 2011; L. Pagani *et al*. 2016; Reyes-Centeno *et al*. 2014). Many of the chrY lineages common in Eurasia coalesce within the last 15 KY (Figure 1, REF I) corresponding to the climate amelioration after the Last Glacial Maximum (LGM). Within the last 4–8 KY numerous novel region-specific clades appear (SI Table S7, SI Figure S11, SI 6 REF I).

Male and female effective population sizes

We used mtDNA and chrY sequences from 320 individuals with known geographic affiliations and reconstructed the temporal changes in regional male and female effective population sizes. The cumulative Bayesian skyline plot (BSP) (SI Figure S4a, REF I) shows increase in both male and female N_e at about 40–60 KYA (Figure 2, REF I), however the female N_e estimates are consistently more than twice as high (SI Figure S6, REF I). Both the chrY and mtDNA show the increase of N_e in the Holocene (from ~11.5 KYA to present), the phenomenon documented before for females (Gignoux *et al*. 2011). Surprisingly, there is a sharp decrease in the chrY N_e around 8–4 KYA without a counterpart in the mtDNA making the female N_e up to 17-fold higher at that time (Figure 2, SI Figure S4B, SI Table S4, SI Figure S5 REF I). This might be explained by either natural selection affecting chrY or culturally driven sex-specific changes in the variance in the offspring number. This reduction is not limited to few haplogroups (SI Figure S3 REF I), therefore selection is not a likely explanation. The drop coincides with the archaeological record characterizing the spread of Neolithic cultures and the demographic changes (Barker 2009). The temporal sequence of the male N_e decline (Supplementary Figure S4B REF I) follows the sequence of the spread of agriculture in the Near East, East Asia, and South Asia earlier, then later in Europe (Bellwood 2004; Fuller 2003). If male reproductive success was even partially culturally inherited, then social changes that increased male variance in the offspring number could explain the drastic drop in the male N_e (Heyer *et al*. 2005).

Population structure also affects N_e. In simple models without competition among demes structure always increases the N_e, but when coupled with unbalanced sampling, population structure can lead BSP to infer false signals of N_e decline under a constant population size model (Heller *et al*. 2013). Such a brief and drastic drop in male N_e is not likely caused by the increased male migration rate, just as the simple models of increase or decrease of the population structure are not sufficient to explain the observed empirical patterns (SI 5, SI Figure S7 REF I). However, when there is competition among demes, an increased variance of expected offspring number among demes could drastically decrease the N_e (Whitlock and Barton 1997). Male-driven conquest,

with an historical example of Mongol expansion, could cause such effect to be solely male-specific (Zerjal *et al*. 2003). Thus, the effect we observe empirically could be due to male-specific cultural inheritance of fitness that causes variance in the number of offspring of males within demes. When combined with sex-biased migration patterns (Destro Bisol *et al*. 2012; Skoglund *et al*. 2014), the male-specific variance among demes is further increased.

5.2 Ancient DNA reveals dual ancestry of Native Americans (REF II)

This study presents a genome of a juvenile individual (MA-1) from Ma'lta Upper Palaeolithic archaeological site in south-central Siberia (Fig1a; Supplementary Information (SI) 1 REF II) with the age 23.9–24.4 KY from the present (SI 2, REF II). The DNA of MA-1 was sequenced to average depth of 1X with good DNA preservation metrics and low contamination rates and overall error rate of 0.27% (SI Table 2, SI 3, 5, 6.1; REF II).

Uniparental lineages of the Mal'ta boy

We confirmed the biological sex of the MA-1, known also as the Mal'ta boy by sequencing his Y chromosome with low depth (1.5× depth for 5.8 million chrY bases). We called the highest frequency bases with base quality 13 or higher in the regions determined for good mapping quality (Wei, Ayub, Chen, *et al*. 2013). To elucidate the relation of the aY and contemporary chrY lineages, we built a phylogenetic tree of 24 publicly available globally selected samples (Drmanac *et al*. 2010) that represent most of the major lineages outside Africa (E, D, G, I, N, Q, R1a, R1b, R2).

The chrY of the Mal'ta boy shared root mutations with all the contemporary hg R lineages, being derived at 19 of them (and additional 17 with missing data), but ancestral in 5 positions (Figure 8; SI 8, Figure SI 5a, REF II). This makes the 24 KY-old Y chromosome an extinct sister lineage to all contemporary hg R lineages. The extant distribution of hg R as a whole is vast, reaching from western Eurasia to south Asia and Altai region in south Siberia (Figure 8; SI 8, Figure SI 5b, REF II). The contemporary hg R sub-clades have many distinct regional frequency patterns, Mal'ta site is located near the eastern fringe of hg R distribution area (Busby *et al*. 2012; Myres *et al*. 2011; Underhill *et al*. 2015) where it represents a minor hg among the contemporary populations (Figure 8; SI Fig 5b, REF II). Except for Altaians and Khakassians with considerably higher proportion of hg R1a (34% and 31% respectively) (Underhill *et al*. 2015), the contemporary chrY pool around the Baikal region in south-central Siberia is dominated mainly by hg N and hg C lineages (Karafet 2002, Malyarchuk 2010, Kharkov 2014, Ilumäe 2016). Mal'ta boy's chrY belongs to a generic branch of the Y chromosome tree that carries West and South Eurasian varieties of haplogroup R and predominantly Siberian and Native American varieties of Q while having no extant lineage detected so far. Thus, his chrY

belongs to the ancestral pool of contemporary western Eurasian paternal lineages and is more distant from the major chrY types prevalent in populations living in south Siberia today. The maternal lineage of MA-1 belongs to mtDNA hg U and similarly to chrY has no affiliation to any contemporary U sub-lineage, thus today also his maternal lineage is rare or extinct (Figure 8; SI 7, SI Figure 4a, REF II). Currently, mtDNA hg U sub-lineages cover a large area encompassing North Africa, Middle East, South and Central Asia, western Siberia and Europe (Figure 8; SI Figure 4b REF II). Notably, mtDNA hg U has been found at high frequency (> 80%) in ancient hunter-gatherers from the Upper Palaeolithic and Mesolithic Europe (Bramanti *et al*. 2009; Günther *et al*. 2018; Jones *et al*. 2017; Malmström *et al*. 2009), so the maternal line of Mal'ta connects pre-agricultural Europe to Upper-Palaeolithic Siberia.

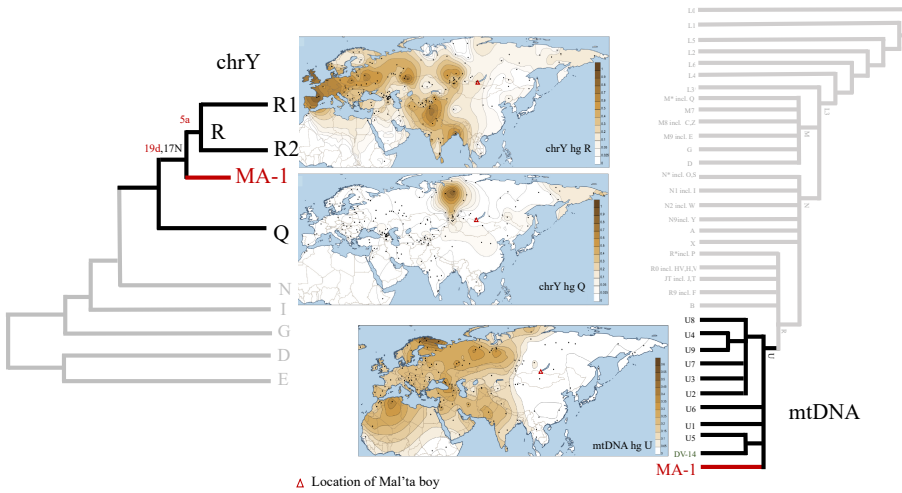

Figure 8. MA-1 paternal and maternal lineages placed on the respective global phylogenetic trees, and the extant Eurasian distributions of chrY hgs R and Q, and mtDNA hg U. The schematic phylogeny (on the left) of chrY sequences from MA-1 and 24 contemporary global samples (Drmanac *et al*. 2010) is based on the unrooted NJ tree (SI 8, Figure SI 5a, REF II). Red numbers show the number of diagnostic positions in the dataset where MA-1 was either in derived (d) or ancestral (a) state, N denotes the number of positions with missing data in MA-1. On the right: schematic global mtDNA tree drawn manually from PhyloTree.org (www.phylotree.org), MA-1 mtDNA sequence belongs to hg U, but does not belong to any extant hg U sub-groups. Middle panels: Extant Eurasian distribution of chrY hgs R and Q (SI8, Figure SI 5b, REF II), and mtDNA hg U (SI7, Figure SI 4b, REF II). Approximate location of MA-1 excavation site shown with red triangle. Figure adapted with permission from Figure SI8, SI5a, SI5b (REF II), Springer Nature.

The dual genomic ancestry of Native Americans and genetic continuity in southern Siberia

Based on the whole genomic signature of MA-1 along the first two components of the PCA in the context of contemporary worldwide populations (Figure 1b, SI10, REF II), the ancient individual is intermediate between modern western Eurasians and Native Americans, but distant from east Asians. The amount of shared genetic history between MA-1 and each of the tested 147 contemporary non-African populations (SI 14.2 REF II) showed that populations from two regions have the greatest genetic affinity to MA-1, the Americas and northeast Europe plus northwest Siberia. Within Europe and Asia the shared drift with MA-1 shows a north-to-south latitudinal cline (Figure 1c, SI Figure 21 and 22, REF II). Most populations in south-central Siberia today show no genetic affinity to MA-1, implying substantial gene flow into the region after the LGM most probably from east Asian sources (SI 9.13, REF II).

On the tree (SI 11 REF II) built of whole genomes of modern individuals from worldwide populations, of the Denisovan individual (Meyer *et al*. 2012), and of the Mal'ta boy, the genome of the Mal'ta boy is placed on the branch basal to all western Eurasians (SI Fig 12 REF II). Bootstrap support for the directional gene flow from MA-1 to Karitiana, a native American population, was 99% (Fig. 2 and SI Fig. 12 and 13, REF II). The *D*-statistic tests on proposed tree-like histories (Green *et al*. 2010) confirm that the genome of the Mal'ta boy shares more recent ancestry with the western Eurasian branch after the split of Europeans and east Asians (SI 13, Supplementary Table 13 REF II). The allele-frequency based *D*-statistic tests (Patterson *et al*. 2012) show that 48 Native American populations of entirely First American ancestry (Reich *et al*. 2012) are equidistant from MA-1 so the admixture event had to occur before the diversification of the First American gene pool (Fig. 3a, SI, section 14.4 and S Fig. 24, REF II). The shared genetic material between these two groups is most likely explained by the gene flow from MA-1 lineage to the ancestors of Native Americans that must have taken place after the ancestors of Native Americans had split from the ancestors of East Asians. Since time estimates for this divergence (Gutenkunst *et al*. 2009; Wall *et al*. 2011) are younger than the age of MA-1, the gene flow in the other direction – from Native American ancestors to MA-1 is not likely.

Thus, the genome of the ancient Mal'ta boy revealed that contemporary Native Americans have mixed origins, resulting from admixture between peoples related to modern-day East Asians and western Eurasians. People moving in from northeast Asia, after their initial settlement of Beringia immediately before and after the LGM, dispersed eventually to the Americas (Goebel 1999). Majority of the contemporary Native American paternal lineages belong to hg Q (Lell *et al*. 2002; Roewer *et al*. 2013) which is the sister lineage to hg R, including MA-1 lineage, making Native American chrY phylogenetically more distant from the East Asian lineages and closer to West Eurasians. Within hgQ the Native American chrY are closely related to paternal lineages among contemporary populations in southern Altai (Matthew C. Dulik *et al*. 2012),

whereas the majority of the Native American maternal lineages are nested within East Asian ones.

To test for the genetic continuity within southern Siberia around LGM another ancient genome from an individual AG-2 from Afontova Gora site dated to post-LGM (17,075–16,750) was sequenced. Though much poorer in sequence quality (SI , REF II), the genome of AG-2 showed close affinities with MA-1 (Fig. 1a, SI 2, 5, 15, SI Fig.29; REF II).

In conclusion, the analyses of the genomic data of the remains of Mal'ta boy have revealed that the Siberian Upper Palaeolithic population has a shared ancestry with both western Eurasians and Native Americans, the latter showing signatures of dual ancestry from ancestors of western Eurasians and ancestors of East Asians. Contrary to the whole genome approach which tackles many independent loci and shows the shared ancestry of MA-1 and Native Americans, the uniparental markers show that neither the paternal nor maternal lineage of MA-1, even if ever crossed Beringia, has survived among the contemporary Native American populations. Also, both the maternal and paternal lineages of the Mal'ta boy are not typical or are absent in populations living today in regions surrounding the Mal'ta site. Thus, the uniparental lineages of MA-1 do not support continuity between the Upper Palaeolithic populations and those living in the region today, and demonstrate the fluidity of the boundaries between the 'western' and 'eastern' types of chrY throughout time.

5.3 Paternal haplogroup H1a1 in India, implications for the demographic history of European Roma (REF III)

Among South Asian autochthonous paternal lineages H-M69 accounts on the average for 22% in different populations (Chaubey *et al*. 2011; Debnath *et al*. 2011; Kumar *et al*. 2007; Sengupta *et al*. 2006; Trivedi *et al*. 2008). The relatively deep coalescence and its low frequency elsewhere suggests that this lineage has its origin in South Asia sometime prior to the Last Glacial Maximum (LGM). The geographic dispersal of hg H-M69 and its offshoots is concentrated mostly to South Asia with rapidly declining frequencies away from it (Trivedi *et al*. 2008). Except for one branch, H1a1a-M82, that is found at higher frequencies also in Europe where, with much reduced diversity, it is the South Asian paternal founder lineage shared by all European Roma groups. The origin of European Roma has been traced to India by linguistic, anthropological and genetic studies and northwestern regions of India are the most often proposed source areas for the founder population (Gresham *et al*. 2001; Matras 2005; Melegh *et al*. 2017; Mendizabal *et al*. 2012; Moorjani *et al*. 2013). This study aims to first provide a comprehensive phylogeographic distribution of hg H1a1a-M82 within India by analysing STR diversity in a database of 7,000 chrY samples from 214 populations, and then seek for the contemporary Indian populations paternally closest to the European Roma who carry the H1a1a-M82 lineage.

Haplogroup H1a1a-M82 in South Asia

Hg H1a1a-M82 is one of the frequent paternal lineages in South Asia (Table S4 REF III). More than 12% of all chrY lineages in South Asia belong to hg H1a1a-M82, with the highest incidence of 20% among southern Indian and lowest 0.2% among northeast Indian populations (Table S1 and S3, REF III). The STR variance and expansion ages reflect the frequency distribution of H1a1-M82 within India, with older ages and highest diversity in the southern areas, decreasing towards northern regions (Figure 1, Table 1, REF III). The expansion times (overall TMRCA ~22 KYA; regional times from about 24–16 KYA) with the associated mean pair-wise difference and haplotype diversity suggest that hg H1a1a-M82 originated in the South Asian pre-LGM paternal gene pool. When comparing populations grouped by language family (Table S1, REF III), the expansion times and variance were highest among the Dravidian speakers (25 ± 3.5 KYA, 0.59) followed by the Austroasiatic speakers (22.4 ± 3.3 KYA, 0.52) and lowest among the Indo-European speakers (19 ± 2.2 KYA, 0.47).

Haplogroup H1a1a-M82 and founders of the European Roma population

Hg H1a1a-M82 among different European Roma populations is at highly reduced genetic diversity relative to that among the South Asian populations. This genetic remnant of South Asian origin supports the hypotheses of a relatively recent migration of a small founding population from India. The average STR-based age estimate of Roma founder lineages splitting from Indian founders is ~1,400 ± 700 YA (Table 3 REF III), which largely agrees with other split-time estimates between the founders of Roma and the South Asian populations (Bánfai *et al*. 2018; Mendizabal *et al*. 2012; Moorjani *et al*. 2013; Morar *et al*. 2004). Analyses of STR networks, pairwise genetic distances, AMOVA results, genetic diversity, and the short distance to the Roma modal haplotype suggests that the paternal lineages of northwestern Indian scheduled caste and tribal populations are the closest ones to the European Roma (Table 1, 2, 4, Figure 2 and Figure 3, REF III). It has been hypothesised that D.oma, the indigenous nomadic people of India, were the ancestors of the European Roma (Grierson 1888). D.oma ranged from Punjab to the lower course of Brahmaputra and many of them remained outcasts or tribals, whereas some were assimilated to the lower castes of Indo-European speakers (van Driem 2001). The results of this study on chrY hg H1a1a STR-diversity and phylogeography within India suggest that only the northwestern part of the diffuse widespread range of the ancestral population of D.oma could have been the source of the founding population of the European Roma. Based on the STR-diversity, the contemporary European Roma hg H1a1a-M82 paternal lineages display the closest affinity to the Y chromosomes carried by contemporary northwestern Indian scheduled caste and tribal populations.

5.4 Sequencing Y chromosomes in the detailed genealogical context – haplogroup R1a lineages among Ashkenazi Levites (REF IV)

Most historic records suggest two main routes of Jewish migration to Europe: through northern Africa towards contemporary Spain, founding the Sephardic Jewry and via Italy towards Rhine Valley, establishing the Ashkenazi Jewry (Fig1 REF IV, (Ben-Sasson 1976; De Lange 1984)). The chrY lineage R1a-M582 shows signs of strong founder effect among the Ashkenazi Levites (Behar *et al*. 2003). Contrary to the initial claims of eastern European introgression (Behar *et al*. 2003), the origin of this hg was later shown, based on few resequenced Y chromosomes and STR analyses, to be more likely Near Eastern than European (Rootsi *et al*. 2013). Still, the limited sample size in the latter study did not provide a comprehensive dated phylogeny which would allow a detailed overview of the origin and spread of the haplogroup and compare with other chrY hgs prevalent among Ashkenazi Jews. Therefore, to fill this gap, in this study we compiled 486 captured Y chromosomal sequences from Ashkenazi Levites, from non-Ashkenazi Levites, from non-Levites with affiliations to different Jewish groups, from Near and Middle Eastern and European populations.

Dated phylogeny of R1a-M582 and distribution of the clade

We constructed the entire phylogeny of R1a that holds a diverse variation of samples throughout the Near and Middle East, Anatolia, Caucasus, Europe and the Indian sub-continent. The known most frequent branches are the mainly eastern European sub-clades of R1a1-Z283 (Behar *et al*. 2003; Underhill *et al*. 2015) that stem from an early split within the hg (Fig2, Sup Fig S1, REF IV). Within the previously defined Ashkenazi Levite founder lineage R1a-M582 six SNVs define the branch comprising all chrY sampled from Ashkenazi Levites, non-Ashkenazi Levites, Ashkenazi non-Levites, and non-Jews with known or suspected Ashkenazi origin (Fig. 2 and S Fig. S1, REF IV). This clade, named R1a-Y2619, coalesces ~1,700 (1,334–2,200) YA and has a star-like expansion pattern (Table 1, S Fig. S1, REF IV). The sister clades to R1a-Y2619, sampled from Iranian Azeris, a Kerman, a Yazidi and one sample from Iberia, all coalesce at ~ 3,100 (2,620–3,682) YA. Since all sub-branches of R1a-Y2619 encompass Jewish men affiliating to the Levite status, it is most likely that the initial founder of this lineage was a Levite Jew.

Within the R1a-Y2619 clade there were six men with affiliations to the Horowitz Levite family with known history of migration from Girona in Catalonia to Horovice near Prague, three males have genealogical records of a common ancestor born at 1615 CE (Fig 3a,b, REF IV). The reconstructed molecular phylogeny was consistent with genealogical records (Fig 3c, REF IV), which allowed to use their common ancestral node as an internal calibration point in the dating of the whole R1a tree. It is notable though, that the coalescence age estimates are dependent on the dataset used, since using data

from only Illumina platform, the coalescence age for the R1a-Y2619 clade was ~1,400 YA, which still remains within the range of expansion dates of other Ashkenazi paternal founder lineages.

The exact migratory pathway of R1a-Y2619 lineage to Europe cannot be firmly established, but the coalescence age of R1a-Y2619 and our finding of the presence of all its branches among Ashkenazi Jews tend to favour a single route of entry to Europe as part of the Ashkenazi migration and expansion in Europe. The presence of R1a-Y2619 lineage among non-Ashkenazi Levites could reflect the continuous contacts and movements of people between different Jewish communities, testified by autosomal and mtDNA studies (Atzmon *et al*. 2010; Behar *et al*. 2006, 2010). The current size of Ashkenazi population is around 4 million men (DellaPergola 2015) and the estimated percentage of R1a-M582 among them is 7.9% (Rootsi *et al*. 2013). Here we reclassified all R1a-M582 belonging to the sub-lineage R1a-Y2619, so the number of Ashkenazi men today descending from the relatively recent ancestor can be estimated to be ~300,000, including many of them self-affiliating as Levite.

Coalescence ages of common Ashkenazi haplogroups

To distinguish if the founder effect of R1a-Y2619 among Ashkenazi Levites is specific to that group only, or a part of the general Ashkenazi expansion, we studied other chrY lineages shown to be prevalent among the Ashkenazi Jews (Behar *et al*. 2004; Hammer *et al*. 2000, 2009). The coalescence ages of the hgs G-M377 and Q-M242, which both constitute 5% of Ashkenazi paternal lineages, are ~1,200 YA and ~1,600 YA respectively (Table 1, REF IV). Similarly, different hg E sub-lineages (E-Z838, E-PF3780, E-B923 and E-B933) that cumulatively make up ~20% of the Ashkenazi paternal variation, coalesce between ~1,200–1,600 YA (Table 1, Suppl Fig S3, REF IV). Different from these are the signals from hgs T-M70 and R1b-M269. Hg T-M70 is prevalent in the Middle East and the Ashkenazi Jewish T-M70 lineages sequenced here coalesce deep, ~9,950 YA representing ancient Levantine diversity (Table 1, REF IV). Haplogroup R1b-M269, currently prevalent in Western Europe, shows a very different pattern (Sup Fig S7, REF IV) of repeated single introgressions of Western European chrY lineages to the Ashkenazi Jewish gene pool, as reported previously (Adams *et al*. 2008).

The comparison of hg R1a-Y2619 and the most abundant founder lineage of Ashkenazi Cohens within hg J-P58 showed that Y chromosomes of both lineages form a tight cluster. Similarly to the distribution of hg R1a-Y2619 within Ashkenazi and non-Ashkenazi Levites, the Cohen J1-P58 is shared only with non-Ashkenazi Cohens and its sister-clades have Middle Eastern origin (Supplemental Figure S2, REF IV). On the other hand, the coalescence age of the Cohen J1-P58 is further back in the pre-Diaspora period at about 2,500 YA (Table 1, REF IV).

In conclusion, the comprehensive phylogeny based on sequences of six million bp from chrY shows that all the sampled R1a-582 Jewish men with Ashkenazi ancestry belonged to sub-hg R1a-Y2619 with the coalescence age of

~1,700, which suggests that this hg was most probably part of the initial Ashkenazi expansion from the Middle East to Europe. The age of the expansion time of this hg among Ashkenazi Levites is similar to the expansion times of other hgs prevalent among Ashkenazi Jewish men. Therefore the founder event was most probably part of the general Ashkenazi expansion and not specific to the Levites. Since all sub-branches of the hg have Jewish men affiliating to the Levite status, the most likely initial founder of this lineage was a Levite Jew. The sister branches of R1a-Y2619 are found in individuals from the Middle East, suggesting that the place of origin of the hg can be ascribed to that region.

6. CONCLUSIONS

- The human chrY mutation rate calibrated with data from two ancient individuals is 0.74×10^{-9} per base pair per year and the age of the most recent common ancestor of chrY phylogeny between hg A00 and the rest of the lineages (A2'5) is at about 254 KY.
- There is a 15 KY gap in between the first split of main non-African lineages at about 78–72 KYA and their subsequent differentiation into major hgs now common in Eurasia, America and Oceania at 47–52 KYA, which in turn is followed by fast diversification process.
- The analyses of dynamics of male-specific aspects of N_e revealed a sharp decrease of chrY N_e around 8–4 KYA without a counterpart in the mtDNA data. At the extreme the female N_e was estimated to have been up to 17-fold higher than the male N_e at the same time. This drastic drop in male N_e can be explained by culturally inherited variance of reproductive success among males, potentially related to the accumulation of wealth during Neolithic revolution.
- The paternal and maternal lineages of the 24 KY old MA-1 individual are both not wide-spread among the contemporary populations in the Baikal region, imply closer affinities to the West Eurasians and show population turn-over in the Baikal region. The genome-wide variation of MA-1 derives from a population more closely related to the present-day western Eurasian and Native American than to East Asian and Oceanian populations. On average, Native American genomes are more similar to East than West Eurasian genomes, however their MA-1 affinity suggests that Native Americans have dual ancestry manifesting as a mixture of ancestors of western Eurasians and ancestors of East Asians. Yet, the two uniparental lineages of MA-1 have not been thus far found among the American uniparental gene pool.
- The results based on hg H1a1a STR diversity and phylogeography suggest that paternal lineages of northwestern Indian scheduled caste and tribal populations are the closest ones to the common South Asian founder lineages among European Roma populations.
- Deep sequencing of Y chromosomes from Ashkenazi Levites show that all Ashkenazi Levites who carry hg R1a lineages belong to a tight star-like cluster R1a-Y2619. This cluster has a coalescence age (~1,700 years) which is similar to other Ashkenazi paternal lineages. The geographical origin of the carriers of sister-branches of hg R1a-Y2619 suggest the most likely origin of the haplogroup to be in the Middle East. R1a-Y2619 was thus likely among the initial paternal founders of the Ashkenazi Levites, but its expansion was part of the general population expansion of the Ashkenazi Jews and not a specific event to the Levites.

SUMMARY IN ESTONIAN

Vaade inimese Y kromosoomile – fülogenees, populatsiooni dünaamika ja asutajasündmused

Iga inimese genoomi on demograafilised ja evolutsioonilised protsessid jätnud oma jälje. Tänased tehnoloogiad võimaldavad meil neid jälgi dešifreerida ning neist moodustuvate mustrite süstemaatiline analüüs annab infot nii genoomi kandva üksikisiku, tema rahva, kui ka kogu inimkonna ajaloo paljude aspektide kohta. Inimpopulatsioonide demograafilise ajaloo ja inimese evolutsiooni analüüsimine ja tõlgendamine vahetult geneetilise materjali põhjal sai hoo sisse väikeste, kuid väga oluliste ja iseloomulike genoomi osade uuringutest. Genoomi haploidsed osad – Y-kromosoom ja mitokondri DNA (mtDNA) kajastavad inimese isa- ja emaliinide ajalugu. Populatsiooni ajalugu saab nende abil iseloomustada siis, kui uuringusse on kaasatud palju indiviide, et tagada võimalikult paljude selles populatsioonis esinevate liinide esindatus. Samas suudame me täna 'lugeda' andmeid kogu inimese genoomist järjendades selles leiduvat infot kas täielikult või siis genotüpiseerides sadu tuhandeid, isegi miljoneid üksikuid punktmutatsioone (SNV – *single nucleotide variant* ingl. k.). Erinevalt haploidsetest genoomi piirkondadest, mis päranduvad mööda ema- ja isaliine edasi peaaegu muutumatute tervikutena, segatakse kogu ülejäänud genoomis emalt ja isalt päranduvad osad iga põlvkonna alguses rekombinatsiooni käigus. Seetõttu on ühe indiviidi erinevad genoomi osad erineva geneetilise ajalooga. Analüüsides üle kogu genoomi suurt hulka sõltumatuid testpunkte saab demograafilisi protsesse rekonstrueerida ka uuringusse vähemal hulgal indiviide kaasates. Eriti väärtuslik on selline info arheoloogilistest leidudest saadud geneetilise info dešifreerimisel e. vana-DNA uuringutes, mille puhul esindavad populatsiooni paratamatult vaid need üksikud indiviidid, kelle säilmeid on õnnestunud leida ja analüüsida. Analüüsides erineva pärandumismustriga genoomi osasid nii tänapäeval elavate indiviidide kui ka täpselt dateeritud arheoloogiliste leidude geneetilise materjali alusel, saab oluliselt paremini iseloomustada nii kaasaegsete populatsioonide geneetilist struktuuri kui ka leida uusi tahke populatsioonide ajaloos. Sellesse raamistikku asetuvad ka käesoleva doktoritöö aluseks olevad neli uuringut, kus keskendutakse läbivalt Y-kromosoomide varieeruvuse erinevatele küsimustele.

Y-kromosoom hõlmab keerulise ehitusega kõige pikemat ühtse tervikuna edasi kanduvat haploidset osa genoomist. Täna kasutatav DNA-järjestuse määramise meetod – lühikeste lugemite sekveneerimine (LLS), on avanud võimaluse suure hulga Y-kromosoomi järjestuste analüüsimiseks, kuid endiselt kaasnevad sellega tehnilised väljakutsed. Käesoleva doktoritöö üheks fookuseks on 456 maailma eri piirkondadest pärit mehe Y-kromosoomi resekveneerimine. Uuringusse kaasati võimalikult palju peamisi laialt levinud isaliine eesmärgiga määrata esmalt ära üle kogu Y-kromosoomi piirkonnad, mida LLS meetodiga uurida saab, seejärel rekonstrueerida saadud andmestikust detailne dateeritud

fülogeneesipuu ning võrrelda isa- ja emaliinide efektiivse populatsiooni suurust läbi aja.

Vana-DNA uuringute iseärasuseks on, et uuritavat algmaterjali on väga vähe, kuid ka madala katvusega sekveneeritud iidne genoom annab uudse ning olulise geneetilise momentvõtte minevikust. See võimaldab iseloomustada mineviku populatsioonide geneetilist struktuuri ning avada seoseid kaasaegsete populatsioonide põlvnemise ning dünaamika kohta läbi aja. Käesolev doktoritöö hõlmab uuringut, mille käigus sekveneeriti Lõuna-Siberist pärit 24000 aasta vanuse ülempaleoliitilise Baikali-äärse Malta kultuuri esindaja (MA-1) genoom, millest analüüsiti lisaks ülegenoomsetele markeritele ja emaliinile ka isaliin.

Asutajasündmused, mil uuele inimgrupile paneb aluse tavaliselt väike alamhulk mingist algpopulatsioonist, jätavad tekkinud populatsiooni järeltulevate põlvede geneetilisse pärandisse iseloomuliku jälje. Ühe ema- või isaliini üleesindatus ning selle vähene varieeruvus populatsiooni sees, mis väljendub nn. tähekujulises fülogeneesipuus viitavad võimalikule asutajasündmusele populatsiooni ajaloos. Käesolevas töös analüüsitakse kahte populatsiooni, mille isaliinides on juba varem näidatud selliste jälgede olemasolu.

Lõuna-Aasia ja India on kultuuriliselt ning geneetiliselt väga mitmekesised, tänaste populatsioonide geneetilise struktuuri taga on erinevad tekkelood, mille iseloomustamine alles kestab. Lõuna-Aasiale omase isaliini, haplogrupi (hg) H1a1a-M82 fülogeograafiat on seni vähe iseloomustatud. See isaliin esineb ka kõigis Euroopa romide populatsioonides ja on üks nende peamisi asutajaliine ning ühtlasi märgiks nende India päritolust. Lingvistiliste ja geneetiliste andmete põhjal on romide algkoduks pakutud peamiselt Loode-India piirkondi, kuid isaliinide algpäritolu täpsemaks lokaliseerimiseks puudus seni põhjalik hg H1a1a-M82 India-sisese leviku analüüs. Käesolevasse doktoritöösse kaasatud uurimus täidab selle lünga, analüüsides hg H1a1a-M82 levikut ja mitmekesisust senisest oluliselt suuremas India isaliinide andmestikus ühes Euroopa romide isaliinidega.

Tänapäeva juutide esivanemad, vanad heebrealased, olid semiidikeelne rahvas, kes olulise perioodi oma pikast ajaloost elas Lähis-Idas Levandi piirkonnas. Ajaloo keerdkäikude tõttu on juudid rahvana tänini killustatud paljude geograafiliste piirkondade vahel, kuid on olulises osas säilitanud oma usulised ja kultuurilised traditsioonid. Juudid elasid ning elavad peamiselt Lähis-Idas, Euroopas, Aafrikas, kuid ka Indias ja mujal, sh Lõuna- ning, eriti tänapäeval, Põhja-Ameerikas. Aškenaasi juutide kogukond tekkis Lääne-Euroopas Reini jõe kallastel 8.–10. sajandil ja levis hiljem sealt edasi Kesk- ja Ida-Euroopasse. Tänapäeval elab maailma eri piirkondades kokku umbes 10 miljonit aškenaasi juuti. Kõigi juudikogukondade seas on siiani eristatavad kahe traditsioonilise preestrite klassi – leviitide ja kohenite järeltulijad. Sarnaselt Y-kromosoomile päranduvad leviidi ja koheni nimetused isalt pojale, pakkudes seega head võimalust võrrelda kultuuriliste ja bioloogiliste tunnuste koospärandumist. Aškenaasi leviitide seas on tuvastatud väga kõrge sagedusega isaliin R1a, mida esmalt peeti idaeurooplastega segunemise märgiks, kuid hilisemad uuringud viitasid, et tegu on Lähis-Ida päritolu asutajaliiniga. Lahtiseks jäi aga selle isaliini lahkne-

mise aeg teistest R1a alamharudest. Polnud ka teada, kas liini sage esinemine aškenaasi leviitide seas on seotud vaid sellele grupile omaste sündmustega, või peegeldab pigem üldist aškenaasi populatsiooni laienemist. Lähtudes nendest küsimustest sekveneeriti käesoleva töö käigus 486 peamiselt aškenaasi leviitide, kuid ka teiste aškenaasi ja mitte-aškenaasi juutide Y-kromosoomi ja analüüsiti neid teiste populatsioonide kontekstis.

Doktoritöö peamised tulemused on kokku võetud järeldustes:

- Inimese isaliinide fülogeneetilise puu viimane ühine eellane elas Aafrikas umbes 250 tuhat aastat tagasi (at) (95% usalduspiiridega 192–307 at), peamised mitte-aafrika liinid lahknevad omavahel lühikese ajaperioodi jooksul 47–52 tuhat at. Paljude tänapäevani säilinud liinide ekspansioon toimus viimase 15 tuhande aasta sees.
- Meeste efektiivse populatsiooni suuruses (N_e) toimus järsk langus 4–8 tuhat at, mida emaliinides ei esine. Suurima madalseisu ajal on meeste ja naiste N_e vahe 17-kordne. Erinevad lihtsad populatsioonistruktuuri mudelid seda ei seleta. Geograafiliselt levib N_e langus samas järgnevuses kui neoliitilised muutused kultuuris, mis seisnesid inimeste elukorralduse muutuses ülemineku küttimiselt-koriluselt põlluharimisele. Seega võib "pudelikael" (meeste reproduktiivse edukuse hajuvuse suurenemine deemide vahel ja sees) olla tingitud meeste reproduktiivset käitumist kujundanud kultuurilistest teguritest.
- Lõuna-Siberist pärit Baikali-äärse 24000 aasta vanuse ülempaleoliitilise Malta kultuuri esindaja MA-1 ema- ja isaliin ei ole tüüpilised selles piirkonnas tänapäeval laialt levinud liinid. Y-kromosoomi hg R, mille basaalsesse harusse MA-1 isaliin kuulub, on tänapäeval levinud Lääne-Euraasiast Lõuna-Aasiani. MA-1 emaliin, mtDNA hg U, mida on leitud ka Euroopa ülempaleoliitiliste küttide-korilaste seast, on samuti üks tänapäeval Baikali ümbruses vähem levinud liinidest. Ka ülegenoomsete markerite analüüs näitab, et tänapäeva Siberi kesk- ja lõunaosas elavatel populatsioonidel ei ole MA-1-ga olulist geneetilist sarnasust, mis viitab suurele jääajajärgsele inimeste sissevoolule sellesse piirkonda Ida-Aasiast.
- Seni ainus geneetiliselt analüüsitud Lõuna-Siberi ülempaleoliitilise populatsiooni esindajal MA-1 jagab ülegenoomseid mustreid nii Lääne-Euraasia populatsioonidega kui Ameerika põlisasukatega, mis viitab ühisele geneetilisele ajaloole. Need leiud lubavad püstitada hüpoteesi, et tänapäeva Ameerika põliselanike eellased kujunesid Lääne-Euraasia ja Ida-Aasia populatsioonide eellaste segunemise tulemusena. Segunemine pidi toimuma enne tänapäeva Ameerika põlisasukate jagunemist erinevateks populatsioonideks.
- Y-kromosoomi haplogrupp H1a1-M82 esineb 12% Lõuna-Aasia meestel, kõrgeim selle liini sagedus on Lõuna-Indias (20%), madalaim aga India kirdeosas (0.2%). Lühikeste tandeemsete korduste järgi arvutatult algas selle hg ekspansioon ~22000 at. Kõige suurem varieeruvus esineb Lõuna-Indiast pärit meestel. Hg H1a1a-M82 madal liinisisene mitmekesisus erinevate Euroopa romide rahvakildude seas kinnitab suhteliselt hiljutist väikese

asutajapopulatsiooni väljarännet Indiast. Samuti romide isaliinide lähedust Loode- ja Põhja-Indiast pärit sotsiaalselt alla surutud kastisüsteemi kuuluvate rahvastega ning hõimude esindajatega.

- Aškenaasi juutide Y-kromosoomide kõrge katvusega sekveneerimine näitas, et kõigi aškenaasi leviidite R1a haplogrupi Y-kromosoomid moodustavad tähekujulise alamklaadi R1a-Y2619, mis koalestseerub, sarnaselt teistele aškenaasi isaliinidele, keskmiselt 1740 aastat tagasi. Kinnitust leidis, et R1a-Y2619 on kõige tõenäolisemalt pärit Lähis-Idast ning võis kuuluda aškenaasi leviitide esmaste asutajaliinide hulka. Haplogrupp levis aškenaasi juutide populatsiooni üldise ekspansiooni käigus ega olnud seotud leviitidele ainuomaste populatsioonisündmustega.

REFERENCES

1000 Genomes Project Consortium, Abecasis GR, Altshuler DL, Bentley DR, Chakravarti A, Clark AG, *et al*. 2010. A Map of Human Genome Variation from Population-Scale Sequencing. *Nature* 467(7319):1061–73.

1000 Genomes Project Consortium, McVean GA, Altshuler DM, Durbin RM, Abecasis GRGR, Bentley DR, *et al*. 2012. An Integrated Map of Genetic Variation from 1,092 Human Genomes. *Nature* 491(7422):56–65.

Adams SM, Bosch E, Balaresque PL, Ballereau SJ, Lee AC, Arroyo E, *et al*. 2008. The Genetic Legacy of Religious Diversity and Intolerance: Paternal Lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *American Journal of Human Genetics* 83(6):725–36.

Al-Zahery N, Pala M, Battaglia V, Grugni V, Hamod MA, Hooshiar 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):288.

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

Ammerman A and Cavalli-Sforza L. 1973. A Population Model for the Difussion of Early Farming in Europe. Pp. 335–58 in *The explanation of cultural change: models in prehistory* , edited by C. Renfrew. London: Duckworth

Anon. 1972. *Encyclopedia Judaica*

Anon. n.d. Katrak SK. Who Are the Parsees? Karachi: Pakistan Herald Press; 1965.

Archidiacono N, Storlazzi CT, Spalluto C, Ricco AS, Marzella R, and Rocchi M. 1998. Evolution of Chromosome Y in Primates. *Chromosoma* 107(4):241–46.

Armitage SJ, Jasim SA, Marks AE, Parker AG, Usik VI, and Uerpmann H-P. 2011. The Southern Route "Out of Africa": Evidence for an Early Expansion of Modern Humans into Arabia. *Science* 453(6016):453–156.

Atzmon G, Hao L, Pe'er I, Velez C, Pearlman A, Palamara PF, *et al*. 2010. Abraham's Children in the Genome Era: Major Jewish Diaspora Populations Comprise Distinct Genetic Clusters with Shared Middle Eastern Ancestry. *The American Journal of Human Genetics* 86(6):850–59.

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(S1):18.

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):e0122968.

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 Forensic Implications. *The American Journal of Human Genetics* 87(3):341–53.
Bánfai Z, Ádám V, Pöstyéni E, Büki G, Czakó M, Miseta A, *et al*. 2018. Revealing the Impact of the Caucasus Region on the Genetic Legacy of Romani People from Genome-Wide Data. *PLoS ONE* 13(9)(e0202890):1–17.
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.
Barešić A and Peričić Salihović M. 2014. Carrier Rates of Four Single-Gene Disorders in Croatian Bayash Roma. *Genetic Testing and Molecular Biomarkers* 18(2):83–87.
Barker G. 2009. *The Agricultural Revolution in Prehistory: Why Did Foragers Become Farmers?* Oxford University Press, USA
Baskovich B, Hiraki S, Upadhyay K, Meyer P, Carmi S, Barzilai N, *et al*. 2016. Expanded Genetic Screening Panel for the Ashkenazi Jewish Population HHS Public Access. *Genet Med* 18(5):522–28.
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):7152.
Batini C and Jobling MA. 2017. Detecting Past Male-Mediated Expansions Using the Y Chromosome. *Human Genetics* 136(5):547–57.
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, Garrigan D, Kaplan ME, Mobasher Z, Rosengarten D, Karafet TM, *et al*. 2004. Contrasting Patterns of Y Chromosome Variation in Ashkenazi Jewish and Host Non-Jewish European Populations. *Human Genetics* 114(4):354–65.
Behar DM, Metspalu E, Kivisild T, Achilli A, Hadid Y, Tzur S, *et al*. 2006. The Matrilineal Ancestry of Ashkenazi Jewry: Portrait of a Recent Founder Event. *The American Journal of Human Genetics* 78(3):487–97.
Behar DM, Metspalu M, Baran Y, Kopelman NM, Yunusbayev B, Gladstein A, *et al*. 2013. No Evidence from Genome-Wide Data of a Khazar Origin for the Ashkenazi Jews. *Human Biology* 85(6):859–900.
Behar DM, Thomas MG, Skorecki K, Hammer MF, Bulygina E, Rosengarten D, *et al*. 2003. Multiple Origins of Ashkenazi Levites: Y Chromosome Evidence for Both Near Eastern and European Ancestries. *The American Journal of Human Genetics* 73(4):768–79.
Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, *et al*. 2008. The Dawn of Human Matrilineal Diversity. *The American Journal of Human Genetics* 82(5):1130–40.
Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, Parik J, *et al*. 2010. The Genome-Wide Structure of the Jewish People. *Nature* 466(7303):238–42.

Beleza S, Gusmão L, Amorim A, Carracedo A, and Salas A. 2005. The Genetic Legacy of Western Bantu Migrations. *Human Genetics* 117(4):366–75.

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.

Bellwood P. 2004. *First Farmers: The Origins of Agricultural Societies*. Wiley

Ben-Sasson HH. 1976. A History of the Jewish People. Harvard University Press

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.

Besaggio D, Fuselli S, Srikummool M, Kampuansai J, Castrì L, Tyler-Smith C, *et al*. 2007. Genetic Variation in Northern Thailand Hill Tribes: Origins and Relationships with Social Structure and Linguistic Differences. *BMC Evolutionary Biology* 7(SUPPL. 2):S12.

Bhowmick BK, Satta Y, and Takahata N. 2007. The Origin and Evolution of Human Ampliconic Gene Families and Ampliconic Structure. *Genome Research* 17(4):441–50.

Bishop C, Guellaen G, Geldwerth D, Fellous M, and Weissenbach J. 1984. Extensive Sequence Homologies between Y and Other Human Chromosomes. *Journal of Molecular Biology* 173(4):403–17.

Bowers WP. 1975. Jewish Communities in Spain in the Time of Paul the Apostle. *The Journal of Theological Studies* 26(2):395–402.

Bowler JM, Johnston H, Olley JM, Prescott JR, Roberts RG, Shawcross W, *et al*. 2003. New Ages for Human Occupation and Climatic Change at Lake Mungo, Australia. *Nature* 421(6925):837–40.

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.

Bray SM, Mulle JG, Dodd AF, Pulver AE, Wooding S, and Warren ST. 2010. Signatures of Founder Effects, Admixture, and Selection in the Ashkenazi Jewish Population. *Proceedings of the National Academy of Sciences* 107(37):16222–27.

Broushaki F, Thomas MG, Link V, López S, van Dorp L, Kirsanow K, *et al*. 2016. Early Neolithic Genomes from the Eastern Fertile Crescent. *Science* 353(6298):499–503.

Browning SR, Browning BL, Zhou Y, Tucci S, and Akey JM. 2018. Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell* 173:53–61.

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.

Burgoyne PS. 1998. The Mammalian Y Chromosome: A New Perspective. *BioEssays* 20(5):363–66.

Busby GBJ, Brisighelli F, Sá Nchez-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.

Cadenas AM, Zhivotovsky LA, Cavalli-Sforza LL, Underhill PA, and Herrera RJ. 2008. Y-Chromosome Diversity Characterizes the Gulf of Oman. *European Journal of Human Genetics* 16(3):374–86.

Cahn C and Guild E. 2008. Recent Migration of Roma in Europe. *Organization for Security and Co-Operation in Europe High Commissioner on National Minorites* (December 2008):1–86.

Cann RL, Stoneking M, and Wilson AC. 1987. Mitochondrial DNA and Human Evolution. *Nature* 325:31–36.

Carmelli D and Cavalli-Sforza LL. 1979. The Genetic Origin of the Jews: A Multivariate Approach. *Human Biology* 51(1):41–61.

Carmi S, Hui KY, Kochav E, Liu X, Xue J, Grady F, *et al*. 2014. Sequencing an Ashkenazi Reference Panel Supports Population-Targeted Personal Genomics and Illuminates Jewish and European Origins. *Nature Communications* 5(45):4835.

Cassidy LM, Martiniano R, Murphy EM, Teasdale MD, Mallory J, Hartwell B, *et al*. 2016. Neolithic and Bronze Age Migration to Ireland and Establishment of the Insular Atlantic Genome. *Proceedings of the National Academy of Sciences* 113(2): 368–73.

Charchar FJ, Svartman M, El-Mogharbel N, Ventura M, Kirby P, Matarazzo MR, *et al*. 2003. Complex Events in the Evolution of the Human Pseudoautosomal Region 2 (PAR2). *Genome Research* 13(2):281–86.

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

Charlesworth B. 1991. Evoluton of Sex Chromosomes. *Science* 251:1030–33.

Chaubey G, Metspalu M, Choi Y, Mägi R, Romero IG, Soares P, *et al*. 2011. Population Genetic Structure in Indian Austroasiatic Speakers: The Role of Landscape Barriers and Sex-Specific Admixture. *Molecular Biology and Evolution* 28(2):1013–24.

Chennakrishnaiah S, Perez D, Gayden T, Rivera L, Regueiro M, and Herrera RJ. 2013. Indigenous and Foreign Y-Chromosomes Characterize the Lingayat and Vokkaliga Populations of Southwest India. *Gene* 526(2):96–106.

Chiaroni J, King RJ, Myres NM, Henn BM, Ducourneau A, Mitchell MJ, *et al*. 2010. The Emergence of Y-Chromosome Haplogroup J1e among Arabic-Speaking Populations. *European Journal of Human Genetics* 18(3):348–53.

Chiaroni J, Underhill PA, and Cavalli-Sforza LL. 2009. Y Chromosome Diversity, Human Expansion, Drift, and Cultural Evolution. *Proceedings of the National Academy of Sciences* 106(48):20174–79.

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

Copeland SR, Sponheimer M, De Ruiter DJ, Lee-Thorp JA, Codron D, Le Roux PJ, *et al*. 2011. Strontium Isotope Evidence for Landscape Use by Early Hominins. *Nature* 474(7349):76–78.

Cordaux R. 2004. The Northeast Indian Passageway: A Barrier or Corridor for Human Migrations? *Molecular Biology and Evolution* 21(8):1525–33.

Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, *et al*. 2014. Origins and Functional Evolution of Y Chromosomes across Mammals. *Nature* 508(7497):488–93.

Costa MD, Pereira JB, Pala M, Fernandes V, Olivieri A, Achilli A, *et al*. 2013. A Substantial Prehistoric European Ancestry amongst Ashkenazi Maternal Lineages. *Nature Communications* 4.

Cotter DJ, Brotman SM, and Wilson Sayres MA. 2016. Genetic Diversity on the Human x Chromosome Does Not Support a Strict Pseudoautosomal Boundary. *Genetics* 203(1):485–92.

Di Cristofaro J, Pennarun E, Mazières S, Myres NM, Lin AA, Temori SA, *et al*. 2013. Afghan Hindu Kush: Where Eurasian Sub-Continent Gene Flows Converge. *PLoS ONE* 8(10).

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, 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, 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.

Cruciani F, Trombetta B, Sellitto D, Massaia A, Destro-Bisol G, Watson E, *et al*. 2010. Human Y Chromosome Haplogroup R-V88: A Paternal Genetic Record of Early Mid Holocene Trans-Saharan Connections and the Spread of Chadic Languages. *European Journal of Human Genetics* 18(7):800–807.

Debnath M, Palanichamy MG, Mitra B, Jin J-Q, Chaudhuri TK, and Zhang Y-P. 2011. Y-Chromosome Haplogroup Diversity in the Sub-Himalayan Terai and Duars Populations of East India. *Journal of Human Genetics* 56(11):765–71.

Delbridge M. 2004. TSPY, the Cadidate Gonadoblastoma Gene on the Human Y Chromosome, Has a Widely Expressed Homologue on the X-Implications for Y Chromosome Evolution. *Chromosome Research* 12(Graves 2002):345–56.

Delbridge ML, Lingenfelter PA, Disteche CM, and Marshall Graves JA. 1999. The Candidate Spermatogenesis Gene RBMY Has a Homologue on the Human X Chromosome. *Nature Genetics* 22(3):223–24.

DellaPergola S. 2015. Berman Jewish DataBank. *American Jewish Year Book* 115:273–364.

Délot EC, Papp JC, Délot EC, Fox M, Grody W, Lee H, *et al*. 2017. Genetics of Disorders of Sex Development: The DSD-TRN Experience. *Endocrinology and Metabolism Clinics of North America* 46(2):519–37.

Destro Bisol G, Capocasa M, and Anagnostou P. 2012. When Gender Matters: New Insights into the Relationships between Social Systems and the Genetic Structure of Human Populations. *Molecular Ecology* 21(20):4917–20.

van Driem G. 2001. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region, Containing an Introduction to the Symbiotic Theory of Language,*. Leiden: Brill

Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, *et al*. 2010. Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. *Science* 327(5961):78–81.

Drummond AJ and Rambaut A. 2007. BEAST: Bayesian Evolutionary Analysis by Sampling Trees. *BMC Evolutionary Biology* 7(1):214.

Drummond AJ, Rambaut A, Shapiro B, and Pybus OG. 2005. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Molecular Biology and Evolution* 22(5):1185–92.

Dulik MC, Owings AC, Gaieski JB, Vilar MG, Andre A, Lennie C, *et al*. 2012. Y-Chromosome Analysis Reveals Genetic Divergence and New Founding Native Lineages in Athapaskan- and Eskimoan-Speaking Populations. *Proceedings of the National Academy of Sciences* 109(22):8471–76.

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.

Efron JM. 2013. Jewish Genetic Origins in the Context of Past Historical and Anthropological Inquiries. *Human Biology* 85(6):901–19.

El-Sibai M, Platt DE, Haber M, Xue Y, Youhanna SC, Wells RS, *et al*. 2009. Geographical Structure of the Y-Chromosomal Genetic Landscape of the Levant: A Coastal-Inland Contrast. *Annals of Human Genetics* 73(6):568–81.

Elhaik E. 2013. The Missing Link of Jewish European Ancestry: Contrasting the Rhineland and the Khazarian Hypotheses. *Genome Biology and Evolution* 5(1):61–74.

Ellegren H. 2004. Microsatellites: Simple Sequences with Complex Evolution. *Nature Reviews Genetics* 5(6):435–45.

Falk R. 2015. Genetic Markers Cannot Determine Jewish Descent. *Frontiers in Genetics* 5(JAN).

Ferri G, Tofanelli S, Alù M, Taglioli L, Radheshi E, Corradini B, *et al*. 2010. Y-STR Variation in Albanian Populations: Implications on the Match Probabilities and the Genetic Legacy of the Minority Claiming an Egyptian Descent. *International Journal of Legal Medicine* 124(5):363–70.

Fiore M, Pillois X, Nurden P, Nurden AT, and Austerlitz F. 2011. Founder Effect and Estimation of the Age of the French Gypsy Mutation Associated with Glanzmann Thrombasthenia in Manouche Families. *European Journal of Human Genetics* 19(9):981–87.

Fishberg M. 1911. *The Jews: A Study of Race and Environment*. New York: Walter Scott

Flegontov P, Kassian A, Thomas MG, Fedchenko V, Changmai P, and Starostin G. 2016. Pitfalls of the Geographic Population Structure (GPS) Approach Applied to Human Genetic History: A Case Study of Ashkenazi Jews. *Genome Biology and Evolution* 8(7):2259–65.

Ford CE and Jones KW. 1959. A Sex-Chromosome Anomaly in a Case of Gonadal Dysgenesis (Turner ' s Syndrome). *Lancet* i(7075):711–13.

Forster P, Harding R, Torroni A, and Bandelt H-J. 1996. Origin and Evolution of Native American MtDNA Variation: A Reappraisal. *American Journal of Human Genetics* 59(4):935–45.

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(6145):565–69.

Fraser A. 1995. *The Gypsies, 2nd Edition*. Wiley-Blackwell

Freije D, Helms C, Watson MS, and Donis-Keller H. 1992. Identification of a Second Pseudoautosomal Region Near the Xq and Yq Telomers. *Science* 258(December):784–86.

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(7523):445–49.

Fu Q, Mittnik A, Johnson PLF, Bos K, Lari M, Bollongino R, *et al*. 2013. A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. *Current Biology* 23(7):553–59.

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.

Fuller DQ. 2003. An Agricultural Perspective on Dravidian Historical Linguistics: Archaeological Crop Packages, Livestock and Dravidian Crop Vocabulary. *Assessing the Languaging/Farming Dispersal Hypothesis* (May 2014):191–213.

Gallego Llorente M, Jones ER, Eriksson A, Siska V, Arthur KW, Arthur JW, *et al*. 2015. Ancient Ethiopian Genome Reveals Extensive Eurasian Admixture in Eastern Africa. *Science* 350(October):6–10.

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

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.

Gignoux CR, Henn BM, and Mountain JL. 2011. Rapid, Global Demographic Expansions after the Origins of Agriculture. *Proceedings of the National Academy of Sciences* 108(15):6044–49.

Goebel T. 1999. Pleistocene Human Colonization of Siberia and Peopling of the Americas: An Ecological Approach. *Evol. Anthropol.* 8.

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

Graves JAM. 2006. Sex Chromosome Specialization and Degeneration in Mammals. *Cell* 124(5):901–14.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, *et al*. 2010. A Draft Sequence of the Neandertal Genome. *Science* 328(5979):710–22.

Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, *et al*. 2001. Origins and Divergence of the Roma (Gypsies). *The American Journal of Human Genetics* 69(6):1314–31.

Grierson GA. 1888. Doms, Jats, and the Origin of the Gypsies. *Journal of Gypsy Lore Society* 1:71–76.

Grugni V, Battaglia V, Hooshiar Kashani B, Parolo S, Al-Zahery N, Achilli A, *et al*. 2012. Ancient Migratory Events in the Middle East: New Clues from the Y-Chromosome Variation of Modern Iranians edited by T. Kivisild. *PLoS ONE* 7(7): e41252.

Gunnarsdóttir ED, Li M, Bauchet M, Finstermeier K, and Stoneking M. 2011. High-Throughput Sequencing of Complete Human MtDNA Genomes from the Philippines. *Genome Research* 21(1):1–11.

Günther T, Malmström H, Svensson EM, Omrak A, Sánchez-Quinto F, Kılınç GM, *et al*. 2018. Population Genomics of Mesolithic Scandinavia: Investigating Early Post-glacial Migration Routes and High-Latitude Adaptation edited by N. Barton. *PLOS Biology* 16(1):e2003703.

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.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD, and Mcvean G. 2009. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. edited by G. McVean. *PLoS Genet* 5(10): e1000695.

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, Doumet-Serhal C, Scheib C, Xue Y, Danecek P, Mezzavilla M, *et al*. 2017. Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences. *American Journal of Human Genetics* 101(2):1–9.

Haber M, Platt DE, Ashrafian Bonab M, Youhanna SC, Soria-Hernanz DF, Martínez-Cruz B, *et al*. 2012. Afghanistan's Ethnic Groups Share a Y-Chromosomal Heritage Structured by Historical Events. *PLoS ONE* 7(3).

Haldane JBS. 1947. The Mutation Rate of the Gene for Haemophilia, and Its SegregationRatios in Males and Females. *Annals of Eugenics* 13:262–71.

Hallast P, Batini C, Zadik D, Delser PM, Jon H, Arroyo-pardo E, *et al*. 2015. 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.

Hallast P and Jobling MA. 2017. The Y Chromosomes of the Great Apes. *Human Genetics* 136(5):511–28.

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

Hammer MF, Behar DM, Karafet TM, Mendez FL, Hallmark B, Erez T, *et al*. 2009. Extended Y Chromosome Haplotypes Resolve Multiple and Unique Lineages of the Jewish Priesthood. *Human Genetics* 126(5):707–17.

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. 15(4).

Hammer MF, Karafet TM, Park H, Keiichi A, Ae O, Harihara S, *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.

Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, *et al*. 2000. Jewish and Middle Eastern Non-Jewish Populations Share a Common Pool of Y-Chromosome Biallelic Haplotypes. *Proceedings of the National Academy of Sciences* 97(12):6769–74.

Hancock I. 2002. *We Are the Romani People*. University of Hertfordshire Press

He J-D, Peng M-S, Quang HH, Dang KP, Trieu AV, Wu S-F, *et al*. 2012. Patrilineal Perspective on the Austronesian Diffusion in Mainland Southeast Asia edited by M. Kayser. *PLoS ONE* 7(5):e36437.

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

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

Heller R, Chikhi L, and Siegismund HR. 2013. The Confounding Effect of Population Structure on Bayesian Skyline Plot Inferences of Demographic History edited by T. Mailund. *PLoS ONE* 8(5):e62992.

Henn BM, Hon L, Macpherson JM, Eriksson N, Saxonov S, Pe'er I, *et al*. 2012. Cryptic Distant Relatives Are Common in Both Isolated and Cosmopolitan Genetic Samples. *PLoS ONE* 7(4).

Herrera KJ, Lowery RK, Hadden L, Calderon S, Chiou C, Yepiskoposyan L, *et al*. 2012. Neolithic Patrilineal Signals Indicate That the Armenian Plateau Was Repopulated by Agriculturalists. *European Journal of Human Genetics* 20(3):313–20.

Heyer E, Chaix R, Pavard S, and Austerlitz F. 2012. Sex-Specific Demographic Behaviours That Shape Human Genomic Variation. *Molecular Ecology* 21(3):597–612.

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.

Heyer E, Sibert A, and Austerlitz F. 2005. Cultural Transmission of Fitness: Genes Take the Fast Lane. *Trends in Genetics* 21(4):234–39.

Higham T, Douka K, Wood R, Ramsey CB, Brock F, Basell L, *et al*. 2014. The Timing and Spatiotemporal Patterning of Neanderthal Disappearance. *Nature* 512(7514): 306–9.

Ho SYW and Shapiro B. 2011. Skyline-Plot Methods for Estimating Demographic History from Nucleotide Sequences. *Molecular Ecology Resources* 11(3):423–34.

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.

Hofreiter M, Paijmans JLA, Goodchild H, Speller CF, Barlow A, Fortes GG, *et al*. 2015. The Future of Ancient DNA: Technical Advances and Conceptual Shifts. *BioEssays* 37(3):284–93.

Holden CJ and Mace R. 2003. Spread of Cattle Led to the Loss of Matrilineal Descent in Africa: A Coevolutionary Analysis. *Proceedings of the Royal Society B: Biological Sciences* 270(1532):2425–33.

Horowitz T. 1928. *Toldot Mishpahat Horowitz*

Hublin J-J, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, *et al*. 2017. New Fossils from Jebel Irhoud, Morocco and the Pan-African Origin of Homo Sapiens. *Nature* 546(7657):289–92.

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, and Page DC. 2012. Sequencing of Rhesus Macaque Y Chromosome Clarifies Origins and Evolution of the DAZ (Deleted in AZoospermia) Genes. *BioEssays* 34(12):1035–44.

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

Ilumäe A-M, 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. *The American Journal of Human Genetics* 99(1):163–73.

Jacobs PA and Strong JA. 1959. A Case of Human Intersexuality Having a Possible XXY Sex-Determining Mechanism. *Nature* 183(4657):302–3.

Järve M, Zhivotovsky LA, Rootsi S, Help H, Rogaev EI, Khusnutdinova EK, *et al*. 2009. Decreased Rate of Evolution in Y Chromosome STR Loci of Increased Size of the Repeat Unit. *PLoS ONE* 4(9).

Jobling, Mark A.; Tyler-Smith C. 2003. The Human Y Chromosome: An Evolutionary Marker Comes of Age. *Nature Genetics* 4:598–612.

Jobling M, Hollox EJ, Hurles ME, Kivisild T, and Tyler-Smith C. 2014. *Human Evolutionary Genetics, Second Edition*. Vol. 1

Jobling MA and Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetic* 4(8):598–612.

Jobling MA and Tyler-Smith C. 2017. Human Y-Chromosome Variation in the Genome-Sequencing Era. *Nature Reviews Genetics* 18:489–97.

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* 6:8912.

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.

Jordan FM, Gray RD, Greenhill SJ, and Mace R. 2009. Matrilocal Residence Is Ancestral in Austronesian Societies. *Proceedings of the Royal Society B: Biological Sciences* 276(1664):1957–64.

Jota MS, Lacerda DR, Sandoval JR, Vieira PPR, Ohasi D, Santos-Júnior JE, *et al*. 2016. New Native South American Y Chromosome Lineages. *Journal of Human Genetics* 61(7):593–603.

Kalaydjieva L, Calafell F, Jobling MA, Angelicheva D, De Knijff P, Rosser ZH, *et al*. 2001. Patterns of Inter- and Intra-Group Genetic Diversity in the Vlax Roma as Revealed by Y Chromosome and Mitochondrial DNA Lineages. *European Journal of Human Genetics* 9(2):97–104.

Kalaydjieva L, Gresham D, and Calafell F. 2001. Genetic Studies of the Roma (Gypsies): A Review. *BMC Medical Genetics* 2:5.

Kalaydjieva L, Morar B, Chaix R, and Tang H. 2005. A Newly Discovered Founder Population: The Roma/Gypsies. *BioEssays* 27(10):1084–94.

Kalaydjieva L, Perez-Lezaun A, Angelicheva D, Onengut S, Dye D, Bosshard NU, *et al*. 1999. A Founder Mutation in the GK1 Gene Is Responsible for Galactokinase Deficiency in Roma (Gypsies). *American Journal of Human Genetics* 65(5):1299–1307.

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.

Karlin S, Kenett R, and Bonné-Tamir B. 1979. Analysis of Biochemical Genetic Data on Jewish Populations: II. Results and Interpretations of Heterogeneity Indices and Distance Measures with Respect to Standards. *American Journal of Human Genetics* 31(3):341–65.

Karmin M, Saag L, Vicente M, Wilson Sayres MA, Järve M, Talas UG, *et al*. 2015. A Recent Bottleneck of Y Chromosome Diversity Coincides with a Global Change in Culture. *Genome Research* 25(4):459–66.

Kashimada K and Koopman P. 2010. Sry: The Master Switch in Mammalian Sex Determination. *Development* 137(23):3921–30.
Katsura Y, Iwase M, and Satta Y. 2012. Evolution of Genomic Structures on Mammalian Sex Chromosomes. *Current Genomics* 13(2):115–23.
Kayser M. 2010. The Human Genetic History of Oceania: Near and Remote Views of Dispersal. *Current Biology* 20(4):R194–201.
Kayser M. 2017. Forensic Use of Y-Chromosome DNA: A General Overview. *Human Genetics* 136(5):621–35.
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, Brauer S, Weiss G, Schiefenhövel W, Underhill PA, and Stoneking M. 2001. Independent Histories of Human Y Chromosomes from Melanesia and Australia. *American Journal of Human Genetics* 68(1):173–90.
Kayser M, Choi Y, Van Oven M, Mona S, Brauer S, Trent RJ, *et al*. 2008. The Impact of the Austronesian Expansion: Evidence from MtDNA and Y Chromosome Diversity in the Admiralty Islands of Melanesia. *Molecular Biology and Evolution* 25(7):1362–74.
Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, *et al*. 2004. A Comprehensive Survey of Human Y-Chromosomal Microsatellites. *The American Journal of Human Genetics* 74(6):1183–97.
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.
Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, *et al*. 2000. Characteristics and Frequency of Germline Mutations at Microsatellite Loci from the Human Y Chromosome, as Revealed by Direct Observation in Father/Son Pairs. *American Journal of Human Genetics* 66(5):1580–88.
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* 3:698.
Kemp BM, Malhi RS, McDonough J, Bolnick DA, Eshleman JA, Rickards O, *et al*. 2007. Genetic Analysis of Early Holocene Skeletal Remains from Alaska and Its Implications for the Settlement of the Americas. *American Journal of Physical Anthropology* 132(4):605–21.
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. *The American Journal of Human Genetics* 72(2):313–32.
Klarić IM, Salihović MP, Lauc LB, Zhivotovsky LA, Rootsi S, and Janićijević 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.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, *et al*. 2012. Rate of de Novo Mutations and the Importance of Father-s Age to Disease Risk. *Nature* 488(7412):471–75.

Kopelman NM, Stone L, Wang C, Gefel D, Feldman MW, Hillel J, *et al*. 2009. Genomic Microsatellites Identify Shared Jewish Ancestry Intermediate between Middle Eastern and European Populations. *BMC Genetics* 10:80.

Kovacevic L, Tambets K, Ilumäe A-M, Kushniarevich A, Yunusbayev B, Solnik A, *et al*. 2014. Standing at the Gateway to Europe – the Genetic Structure of Western Balkan Populations Based on Autosomal and Haploid Markers. *PloS One* 9(8): e105090.

Kumar V, Langstieh BT, Madhavi K V., Naidu VM, Singh HP, Biswas S, *et al*. 2006. Global Patterns in Human Mitochondrial DNA and Y-Chromosome Variation Caused by Spatial Instability of the Local Cultural Processes. *PLoS Genetics* 2(4):420–24.

Kumar V, Reddy ANS, Babu JP, Rao TN, Langstieh BT, Thangaraj K, *et al*. 2007. Y-Chromosome Evidence Suggests a Common Paternal Heritage of Austro-Asiatic Populations. *BMC Evolutionary Biology* 7:47.

Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, *et al*. 2001. The AZFc Region of the Y Chromosome Features Massive Palindromes and Uniform Recurrent Deletions in Infertile Men. *Nature Genetics* 29(3):279–86.

Kuroki Y, Toyoda A, Noguchi H, Taylor TD, Itoh T, Kim D-S, *et al*. 2006. Comparative Analysis of Chimpanzee and Human Y Chromosomes Unveils Complex Evolutionary Pathway. *Nature Genetics* 38(2):158–67.

Lacau H, Gayden T, Regueiro M, Chennakrishnaiah S, Bukhari A, Underhill PA, *et al*. 2012. Afghanistan from a Y-Chromosome Perspective. *European Journal of Human Genetics* 20(10):1063–70.

Lahn BT and Page DC. 1997. Functional Coherence of the Human Y Chromosome. *Science* 278(5338):675–80.

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

Lahn BT and Page DC. 2000. A Human Sex-Chromosomal Gene Family Expressed in Male Germ Cells and Encoding Variably Charged Proteins. *Human Molecular Genetics* 9(2):311–19.

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

Lalueza-Fox C, Rosas A, Estalrrich A, Gigli E, Campos PF, Garcia-Tabernero A, *et al*. 2011. Genetic Evidence for Patrilocal Mating Behavior among Neandertal Groups. *Proceedings of the National Academy of Sciences* 108(1):250–53.

Lam HYK, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M, *et al*. 2012. Performance Comparison of Whole-Genome Sequencing Platforms. *Nature Biotechnology* 30(6):562–562.

De Lange N. 1984. Atlas of the Jewish World. Phaidon Press

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* 2(7518):409–13.

Lazaridis IP, 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):419–424.

Lee AC, Kamalam A, Adams SM, and Jobling MA. 2004. Molecular Evidence for Absence of Y-Linkage of the Hairy Ears Trait. *European Journal of Human Genetics* 12(12):1077–79.

Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, *et al*. 2002. The Dual Origin and Siberian Affinities of Native American Y Chromosomes. *The American Journal of Human Genetics* 70(1):192–206.

Li WH, Yi SJ, and Makova K. 2002. Male-Driven Evolution. *Current Opinion in Genetics & Development* 12(6):650–56.

Link V, Aguilar-Gómez D, Ramírez-Suástegui C, Hurst LD, and Cortez D. 2017. Male Mutation Bias Is the Main Force Shaping Chromosomal Substitution Rates in Monotreme Mammals. *Genome Biology and Evolution* 9(9):2198–2210.

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):13.

Livshits G, Sokal RR, and Kobyliansky E. 1991. Genetic Affinities of Jewish Populations. *American Journal of Human Genetics* 49(1):131–46.

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, Roseman C, *et al*. 2004. The Levant versus the Horn of Africa: Evidence for Bidirectional Corridors of Human Migrations. *The American Journal of Human Genetics* 74(3):532–44.

Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DCY, Nazareth L, *et al*. 2010. Whole-Genome Sequencing in a Patient with Charcot–Marie–Tooth Neuropathy. *New England Journal of Medicine* 362(13):1181–91.

Malaspinas A-S, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, *et al*. 2016. A Genomic History of Aboriginal Australia. *Nature* 538(7624):207–14.

Malmström H, Gilbert MTP, Thomas MG, Brandström M, Storå J, Molnar P, *et al*. 2009. Ancient DNA Reveals Lack of Continuity between Neolithic Hunter-Gatherers and Contemporary Scandinavians. *Current Biology* 19(20):1758–62.

Malyarchuk BA, Grzybowski T, Derenko M V, Czarny J, and Miścicka-Sliwka D. 2006. Mitochondrial DNA Diversity in the Polish Roma. *Annals of Human Genetics* 70(Pt 2):195–206.

Malyarchuk BA, Perkova MA, Derenko M V., Vanecek T, Lazur J, and Gomolcak P. 2008. Mitochondrial DNA Variability in Slovaks, with Application to the Roma Origin. *Annals of Human Genetics* 72(2):228–40.

Martínez-Cruz B, Mendizabal I, Harmant C, De Pablo R, Ioana M, Angelicheva D, *et al*. 2016. Origins, Admixture and Founder Lineages in European Roma. *European Journal of Human Genetics* 24(6):937–43.

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.

Matras Y. 1999. Johann Rüdiger and the Study of Romani in 18th Century Germany. *Journal of the Gypsy Lore Society. Ser. 5* 9(9):89–116.

Matras Y. 2005. Romani: A Linguistic Introduction. P. 308 in. Cambridge University Press

Mazeyrat S, Saut N, Mattei MG, and Mitchell MJ. 1999. RBMY Evolved on the Y Chromosome from a Ubiquitously Transcribed X-Y Identical Gene. *Nature Genetics* 22(3):224–26.

Melegh BI, Banfai Z, Hadzsiev K, Miseta A, and Melegh B. 2017. Refining the South Asian Origin of the Romani People. *BMC Genetics* 18(1).

Mellars P, Gori KC, Carr M, Soares PA, and Richards MB. 2013. Genetic and Archaeological Perspectives on the Initial Modern Human Colonization of Southern Asia. *Proceedings of the National Academy of Sciences* 110(26):10699–704.

Mendez FL, Karafet TM, Krahn T, Ostrer H, Soodyall H, and Hammer MF. 2011. Increased Resolution of Y Chromosome Haplogroup T Defines Relationships among Populations of the Near East, Europe, and Africa. *Human Biology* 83(1):39–53.

Mendez FL, Krahn T, Schrack B, Krahn A-M, 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. *The American Journal of Human Genetics* 98(4):728–34.

Mendizabal I, Lao O, Marigorta UM, Wollstein A, Gusmão L, Ferak V, *et al*. 2012. Reconstructing the Population History of European Romani from Genome-Wide Data. *Current Biology* 22(24):2342–49.

Mendizabal I, Valente C, Gusmão A, Alves C, Gomes V, Goios A, *et al*. 2011. Reconstructing the Indian Origin and Dispersal of the European Roma: A Maternal Genetic Perspective edited by R. C. Fleischer. *PLoS ONE* 6(1):e15988.

Meyer M, Kircher M, Gansauge M-TT, Li H, Racimo F, Mallick S, *et al*. 2012. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* 338(6104):222–26.

Mittnik A, Wang C-C, Pfrengle S, Daubaras M, Zariņa G, Hallgren F, *et al*. 2018. The Genetic Prehistory of the Baltic Sea Region. *Nature Communications* 9(1):442.

Moorjani P, Patterson N, Loh PR, Lipson M, Kisfali P, Melegh BI, *et al*. 2013. Reconstructing Roma History from Genome-Wide Data. *PLoS ONE* 8(3).

Morar B, Gresham D, Angelicheva D, Tournev I, Gooding R, Guergueltcheva V, *et al*. 2004. Mutation History of the Roma/Gypsies. *American Journal of Human Genetics* 75(4):596–609.

Mukerjee R. 1937. Caste and Social Change in India. *American Journal of Sociology* 43(3):377.

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.

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.

Nebel A, Filon D, Brinkmann B, Majumder PP, Faerman M, and Oppenheim A. 2001. The Y Chromosome Pool of Jews as Part of the Genetic Landscape of the Middle East. *The American Journal of Human Genetics* 69(5):1095–1112.

Nebel A, Filon D, Faerman M, Soodyall H, and Oppenheim A. 2005. Y Chromosome Evidence for a Founder Effect in Ashkenazi Jews. *European Journal of Human Genetics* 13(3):388–91.

Noy D. 2013. Jews in the Western Roman Empire in Late Antiquity: Migration, Integration, Separation. *Veleia* 30:169–77.

Ohno S. 1966. *Sex Chromosomes and Sex-Linked Genes*. 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.
Olofsson JK, Pereira V, Børsting C, and Morling N. 2015. Peopling of the North Circumpolar Region – Insights from Y Chromosome STR and SNP Typing of Greenlanders. *PLoS ONE* 10(1):1–14.
Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, and Stoneking M. 2001. Human MtDNA and Y-Chromosome Variation Is Correlated with Matrilocal versus Patrilocal Residence. *Nature Genetics* 29(1):20–21.
Orlando L, Gilbert MTP, and Willerslev E. 2015. Reconstructing Ancient Genomes and Epigenomes. *Nature Reviews Genetics* 16(7):395–408.
Ostrer H. 2001. A Genetic Profile of Contemporary Jewish Populations. *Nature Reviews Genetics* 2(11):891–98.
Ostrer H and Skorecki K. 2013. The Population Genetics of the Jewish People. *Human Genetics* 132(2):119–27.
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.
Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, Gallego Romero I, *et al*. 2012. Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool. *American Journal of Human Genetics* 91(1):83–96.
Pagani L, Lawson DJ, Jagoda E, Mörseburg A, Eriksson A, Mitt M, *et al*. 2016. Genomic Analyses Inform on Migration Events during the Peopling of Eurasia. *Nature* 538(7624):238–42.
Page DC, Harper ME, Love J, and Botstein D. 1984. Occurrence of a Transposition from the X-Chromosome Long Arm to the Y-Chromosome Short Arm during Human Evolution. *Nature* 311(5982):119–23.
Palamara PF, Lencz T, Darvasi A, and Pe'er I. 2012. Length Distributions of Identity by Descent Reveal Fine-Scale Demographic History. *American Journal of Human Genetics* 91(5):809–22.
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, Zalán A, Béres J, Nagy M, and Chang YM. 2011. Genetic Structure of the Paternal Lineage of the Roma People. *American Journal of Physical Anthropology* 145(1):21–29.
Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, *et al*. 2012. Ancient Admixture in Human History. *Genetics* 192(3):1065–93.
Perićić M, Lauc LB, Klarić IM, Rootsi S, Janićijevic B, Rudan I, *et al*. 2005. High-Resolution Phylogenetic Analysis of Southeastern Europe Traces Major Episodes of Paternal Gene Flow Among Slavic Populations. *Molecular Biology and Evolution* 22(10):1964–75.
Pfeifer S. 2016. From Next-Generation Resequencing Reads to a High-Quality Variant Data Set. *Nature Publishing Group* 118:111–24.
Pickrell JK, Patterson N, Loh P-RR, Lipson M, Berger B, Stoneking M, *et al*. 2014. Ancient West Eurasian Ancestry in Southern and Eastern Africa. *Proc Natl Acad Sci U S A* 111(7):2632–37.

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.

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, *et al*. 2014. The Complete Genome Sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43–49.

Pybus OG, Rambaut A, and Harvey PH. 2000. An Integrated Framework for the Inference of Viral Population History from Reconstructed Genealogies. *Genetics* 155(3):1429–37.

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, Guo X, Wang Y, Lohmueller KEKE, Rasmussen S, Albrechtsen A, *et al*. 2011. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* 334(6052):94–98.

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.

Rasmussen M, Sikora M, Albrechtsen A, Korneliussen TS, Moreno-Mayar JV, Poznik GD, *et al*. 2015. The Ancestry and Affiliations of Kennewick Man. *Nature* 523 (7561):455–58.

Rasteiro R and Chikhi LS. 2013. Female and Male Perspectives on the Neolithic Transition in Europe: Clues from Ancient and Modern Genetic Data. *PLoS ONE* 8(4).

Ravid-Amir O and Rosset S. 2010. Maximum Likelihood Estimation of Locus-Specific Mutation Rates in Y-Chromosome Short Tandem Repeats. *Bioinformatics* 26(18): 440–45.

Regueiro M, Cadenas AM, Gayden T, Underhill PA, and Herrera RJ. 2006. Iran: Tricontinental Nexus for Y-Chromosome Driven Migration. *Human Heredity* 61(3): 132–43.

Regueiro M, Rivera L, Chennakrishnaiah S, Popovic B, Andjus S, Milasin J, *et al*. 2012. Ancestral Modal Y-STR Haplotype Shared among Romani and South Indian Populations. *Gene* 504(2):296–302.

Regueiro M, Stanojevic A, Chennakrishnaiah S, Rivera L, Varljen T, Alempijevic D, *et al*. 2011. Divergent Patrilineal Signals in Three Roma Populations. *American Journal of Physical Anthropology* 144(1):80–91.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, *et al*. 2010. Genetic History of an Archaic Hominin Group from Denisova Cave in Siberia. *Nature* 468(7327):1053–60.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, *et al*. 2012. Reconstructing Native American Population History. *Nature* 488(7411):370–74.

Reyes-Centeno H, Ghirotto S, Detroit F, Grimaud-Herve D, Barbujani G, and Harvati K. 2014. Genomic and Cranial Phenotype Data Support Multiple Modern Human

Dispersals from Africa and a Southern Route into Asia. *Proceedings of the National Academy of Sciences* 111(20):7248–53.
Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, *et al*. 2000. Tracing European Founder Lineages in the near Eastern MtDNA Pool. *American Journal of Human Genetics* 67(5):1251–76.
Rios J, Stein E, Shendure J, Hobbs HH, and Cohen JC. 2010. Identification by Whole-Genome Resequencing of Gene Defect Responsible for Severe Hypercholesterolemia. *Human Molecular Genetics* 19(22):4313–18.
Risch N, Tang H, Katzenstein H, and Ekstein J. 2003. Geographic Distribution of Disease Mutations in the Ashkenazi Jewish Population Supports Genetic Drift over Selection. *American Journal of Human Genetics* 72(4):812–22.
Roach JC, Glusman G, Smit AFA, Huff CD, Hubley R, Shannon PT, *et al*. 2010. Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing. *Science* 328(5978):636–39.
Roewer L, Nothnagel M, Gusmão L, Gomes V, González M, Corach D, *et al*. 2013. Continent-Wide Decoupling of Y-Chromosomal Genetic Variation from Language and Geography in Native South Americans. *PLoS Genetics* 9(4).
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(1).
Rootsi S, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, *et al*. 2004. Phylogeography of Y-Chromosome Haplogroup I Reveals Distinct Domains of Prehistoric Gene Flow in Europe. *The American Journal of Human Genetics* 75(1):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, Grafham D V., Coffey AJ, Scherer S, McLay K, Muzny D, *et al*. 2005. The DNA Sequence of the Human X Chromosome. *Nature* 434(March):325–37.
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.
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.
Rüdiger JCC. 1782. *Von Der Sprache Und Herkunft Der Zigeuner Aus Indien. Reprint from: Neuester Zuwachs Der Teutsche, Fremden Und Allgemeinen Sprachkunde in Eigenen Aufsätzen*. Hamburg: Buske
Saag L, Varul L, Scheib CL, Stenderup J, Allentoft ME, Saag L, *et al*. 2017. Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe. *Current Biology* 27(14):2185–2193.e6.
Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, Gaikwad S, *et al*. 2006. A Prehistory of Indian Y Chromosomes: Evaluating Demic Diffusion Scenarios. *Proceedings of the National Academy of Sciences* 103(4):843–48.

Saillard J, Forster P, Lynnerup N, Bandelt H-J, and Nørby S. 2000. MtDNA Variation among Greenland Eskimos: The Edge of the Beringian Expansion. *The American Journal of Human Genetics* 67(3):718–26.

Salaman RN. 1911. HEREDITY AND THE JEW. *THE EUGENICS REVIEW* 3:187–200.

Salihović MP, Barešić A, Klarić IM, Cukrov S, Lauc LB, and Janićijević B. 2011. The Role of the Vlax Roma in Shaping the European Romani Maternal Genetic History. *American Journal of Physical Anthropology* 146(2):262–70.

Sankararaman S, Mallick S, Patterson N, and Reich D. 2016. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. *Current Biology* 26(9):1241–47.

Sargent CA, Briggs H, Chalmers IJ, Lambson B, Walker E, and Affara NA. 1996. The Sequence Organization of Yp/Proximal Xq Homologous Regions of the Human Sex Chromosomes Is Highly Conserved. *Genomics* 32(2):200–209.

Saxena R, Brown LLG, Hawkins T, Alagappan RRK, Skaletsky H, Reeve MMP, *et al*. 1996. The DAZ Gene Cluster on the Human Y Chromosome Arose from an Autosomal Gene That Was Transposed, Repeatedly Amplified and Pruned. *Genetics* 14(3):292–99.

Scerri EML, Thomas MG, Manica A, Gunz P, Stock JT, Stringer C, *et al*. 2018. Did Our Species Evolve in Subdivided Populations across Africa, and Why Does It Matter? *Trends in Ecology and Evolution* 33(8):582–94.

Schiffels S, Haak W, Paajanen P, Llamas B, Popescu E, Loe L, *et al*. 2016. Iron Age and Anglo-Saxon Genomes from East England Reveal British Migration History. *Nature Communications* 7:10408.

Schlebusch CM, Malmström H, Günther T, Sjödin P, Coutinho A, Edlund H, *et al*. 2017. Southern African Ancient Genomes Estimate Modern Human Divergence to 350,000 to 260,000 Years Ago. *Science* 358(6363):652–55.

Scozzari R, Massaia A, D'Atanasio E, Myres NM, Perego UA, Trombetta B, *et al*. 2012. Molecular Dissection of the Basal Clades in the Human Y Chromosome Phylogenetic Tree. *PLoS ONE* 7(11).

Scozzari R, Massaia A, Trombetta B, Bellusci G, Natalie M, 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.

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

Ségurel L, Martínez-Cruz B, Quintana-Murci L, Balaresque P, Georges M, Hegay T, *et al*. 2008. Sex-Specific Genetic Structure and Social Organization in Central Asia: Insights from a Multi-Locus Study edited by M. Przeworski. *PLoS Genetics* 4(9): e1000200.

Seielstad MT, Minch E, and Cavalli-Sforza LL. 1998. Genetic Evidence for a Higher Female Migration Rate in Humans. *Nature Genetics* 20(3):278–80.

Semino O, Magri C, Benuzzi G, Lin AAAA, 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.

Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, and Underhill PA. 2002. Ethiopians and Khoisan Share the Deepest Clades of the Human Y-Chromosome Phylogeny. *The American Journal of Human Genetics* 70(1):265–68.

Sengupta S, Zhivotovsky LA, King R, Mehdi SQQ, Edmonds CA, Chow C-ET, *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. *The American Journal of Human Genetics* 78(2):202–21.

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

Shi H, Zhong H, Peng Y, Dong Y-L, Qi X-B, 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):45.

Singh S, Singh A, 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(1):19157.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, 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, Omrak A, Raghavan M, Valdiosera C, Günther T, *et al*. 2014. Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science* 344(6185):747–50.

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.

Skoglund P, Posth C, Sirak K, Spriggs M, Valentin F, Bedford S, *et al*. 2016. Genomic Insights into the Peopling of the Southwest Pacific. *Nature* 538(7626):510–13.

Skoglund P, Thompson JC, Prendergast ME, Mittnik A, Sirak K, Hajdinjak M, *et al*. 2017. Reconstructing Prehistoric African Population Structure. *Cell* 171(1):59–71.

Skorecki K, Selig S, Blazer S, Bradman R, Bradman N, Waburton PJ, *et al*. 1997. Y Chromosomes of Jewish Priests. *Nature* 385(6611):32–32.

Slon V, Mafessoni F, Vernot B, de Filippo C, Grote S, Viola B, *et al*. 2018. The Genome of the Offspring of a Neanderthal Mother and a Denisovan Father. *Nature* 1.

Slon V, Viola B, Renaud G, Gansauge M-T, Benazzi S, Sawyer S, *et al*. 2017. A Fourth Denisovan Individual. *Science Advances* 3(7):e1700186.

Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt H-J, *et al*. 2010. The Archaeogenetics of Europe. *Current Biology* 20(4):174–83.

Stern C. 1957. The Problem of Complete Y-Linkage in Man. *The American Journal of Human Genetics* 9(3):147–66.

Stoneking M. 1998. Women on the Move. *Nature Genetics* 20(3):219–20.

Stringer C. 2002. Modern Human Origins: Progress and Prospects. *Philosophical Transactions of the Royal Society B: Biological Sciences* 357(1420):563–79.

Sun C, Skaletsky H, Birren B, Devon K, Tang Z, Silber S, *et al*. 1999. An Azoospermic Man with a de Novo Point Mutation in the Y-Chromosomal Gene USP9Y. *Nature Genetics* 23(4):429–32.

Sun S and Heitman J. 2012. Should Y Stay or Should Y Go: The Evolution of Non-Recombining Sex Chromosomes. *BioEssays* 34(11):938–42.

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. *Am. J. Hum. Genet* 74:661–82.

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.

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, *et al*. 2012. Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science* 336(6090):64–69.

Thangaraj K, Singh L, Reddy AG, Rao VR, Sehgal SC, Underhill PA, *et al*. 2003. Genetic Affinities of the Andaman Islanders, a Vanishing Human Population. *Current Biology* 13:86–93.

The Y Chromosome Consortium. 2002. A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome Research* 12(2):339–48.

Thomas MG, Skoreckiad K, Ben-Amid H, Parfitt T, Bradman N, and Goldstein DB. 1998. Origins of Old Testament Priests. *Nature* 394(6689):138–40.

Thomas MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, *et al*. 2002. Founding Mothers of Jewish Communities: Geographically Separated Jewish Groups Were Independently Founded by Very Few Female Ancestors. *The American Journal of Human Genetics* 70(6):1411–20.

Thomson R, Pritchard JK, Shen P, Oefner PJ, and Feldman MW. 2000. Recent Common Ancestry of Human Y Chromosomes: Evidence from DNA Sequence Data. *Pnas* 97(13):7360–65.

Tilford C a, 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.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, *et al*. 2009. The Genetic Structure and History of Africans and African Americans. *Science* 324(5930):1035–44.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, *et al*. 2007. Convergent Adaptation of Human Lactase Persistence in Africa and Europe. *Nature Genetics* 39(1):31–40.

Tofanelli S, Brisighelli F, Anagnostou P, Busby GBJ, Ferri G, Thomas MG, *et al*. 2016. The Greeks in the West: Genetic Signatures of the Hellenic Colonisation in Southern Italy and Sicily. *European Journal of Human Genetics* 24(3):429–36.

Torroni A, Achilli A, Macaulay V, Richards M, and Bandelt HJ. 2006. Harvesting the Fruit of the Human MtDNA Tree. *Trends in Genetics* 22(6):339–45.

Trivedi R, Sahoo S, Singh A, Hima Bindu G, Banerjee J, Tandon M, *et al*. 2008. Genetic Imprints of Pleistocene Origin of Indian Populations: A Comprehensive Phylogeographic Sketch of Indian Y-Chromosomes. *International Journal of Human Genetics* 8(1–2):97–118.

Trombetta B, D'Atanasio E, Massaia A, Ippoliti M, Coppa A, Candilio F, *et al*. 2015. Phylogeographic Refinement and Large Scale Genotyping of Human Y Chromosome Haplogroup E Provide New Insights into the Dispersal of Early Pastoralists in the African Continent. *Genome Biol Evol* 7(7):1940–50.

Trombetta B, D'Atanasio E, Massaia A, Myres NM, Scozzari R, Cruciani F, *et al*. 2015. Regional Differences in the Accumulation of SNPs on the Male-Specific Portion of

the Human y Chromosome Replicate Autosomal Patterns: Implications for Genetic Dating. *PLoS ONE* 10(7):1–18.
Turner RL. 1926. The Position of Romani in Indo-Aryan. *Journal of the Gypsy Lore Society* 3(145–94).
Tyler-Smith C, Taylor L, and Müller U. 1988. Structure of a Hypervariable Tandemly Repeated DNA Sequence on the Short Arm of the Human Y Chromosome. *Journal of Molecular Biology* 203(4):837–48.
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 Revolution.* 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, Lahr MM, Foley RA, *et al*. 2001. The Phylogeography of Y Chromosome Binary Haplotypes and the Origins of Modern Human Populations. *Annals of Human Genetics* 65(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 : EJHG* 23(1):124–31.
Underhill PA, Shen P, Lin A, 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.
Vega Genomic Browser. 2017. Vega Genomic Browser 68. *Http://Vega.Archive.Ensembl.Org/Homo_sapiens/Location/Chromosome?R=Y* (68).
Wall JD, Jiang R, Gignoux C, Chen GK, Eng C, Huntsman S, *et al*. 2011. Genetic Variation in Native Americans, Inferred from Latino SNP and Resequencing Data. *Molecular Biology and Evolution* 28(8):2231–37.
Wall JD, Tang LF, Zerbe B, Kvale MN, Kwok PY, Schaefer C, *et al*. 2014. Estimating Genotype Error Rates from High-Coverage next-Generation Sequence Data. *Genome Research* 24(11):1734–39.
Wall JD and Yoshihara Caldeira Brandt D. 2016. Archaic Admixture in Human History. *Current Opinion in Genetics and Development* 41:93–97.
Wang C-C and Li H. 2013. Inferring Human History in East Asia from Y Chromosomes. *Investigative Genetics* 4(1):11.
Wang XV, Blades N, Ding J, Sultana R, and Parmigiani G. 2012. Estimation of Sequencing Error Rates in Short Reads. *BMC Bioinformatics* 13.
Waters PD, Duffy B, Frost CJ, Delbridge ML, and Graves JAM. 2001. The Human Y Chromosome Derives Largely from a Single Autosomal Region Added to the Sex Chromosomes 80-130 Million Years Ago. *Cytogenetics and Cell Genetics* 92(February 2001):74–79.
Weber JL and Wong C. 1993. Mutation of Human Short Tandem Repeats. *Human Molecular Genetics* 2(8):1123–28.
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* 98(18):10244–49.

Whitlock MC and Barton NH. 1997. The Effective Size of Subdivided Population. *Genetics* 146(1):427–41.

Wilder JA, Mobasher Z, and Hammer MF. 2004. Genetic Evidence for Unequal Effective Population Sizes of Human Females and Males. *Molecular Biology and Evolution* 21(11):2047–57.

Wilkins JF and Marlowe FW. 2006. Sex-Biased Migration in Humans: What Should We Expect from Genetic Data? *BioEssays* 28(3):290–300.

Willems T, Gymrek M, Highnam G, 1000 Genomes Project Consortium T 1000 GP, Mittelman D, and Erlich Y. 2014. The Landscape of Human STR Variation. *Genome Research* 24(11):1894–1904.

Willems T, Gymrek M, Poznik GD, Tyler-Smith C, 1000 Genomes Project Chromosome Y Group T 1000 GPCY, and Erlich Y. 2016. Population-Scale Sequencing Data Enable Precise Estimates of Y-STR Mutation Rates. *American Journal of Human Genetics* 98(5):919–33.

Wilson Sayres MA and Makova KD. 2011. Genome Analyses Substantiate Male Mutation Bias in Many Species. *BioEssays* 33(12):938–45.

Wilson Sayres MA and Makova KD. 2013. Gene Survival and Death on the Human Y Chromosome. *Molecular Biology and Evolution* 30(4):781–87.

Xue J, Lencz T, Darvasi A, Pe'er I, and Carmi S. 2017. The Time and Place of European Admixture in Ashkenazi Jewish History. *PLoS Genetics* 13(4).

Xue Y, Wang Q, Long Q, Ng BL, Swerdlow H, Burton J, *et al*. 2009. Human Y Chromosome Base-Substitution Mutation Rate Measured by Direct Sequencing in a Deep-Rooting Pedigree. 19(17).

Yan S, Wang C-C, Li H, Li S-L, 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* 1964(10):1013–15.

Yunusbayev B, Metspalu M, Jarve M, Kutuev I, Rootsi S, Metspalu E, *et al*. 2012. The Caucasus as an Asymmetric Semipermeable Barrier to Ancient Human Migrations. *Molecular Biology and Evolution* 29(1):359–65.

Zalán A, Béres J, and Pamjav H. 2011. Paternal Genetic History of the Vlax Roma. *Forensic Science International: Genetics* 5(2):109–13.

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):2077.

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, *et al*. 2017. Ensembl 2018. *Nucleic Acids Research* 46(D1):D754–61.

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 RS, Bao W, Zhu S, *et al*. 2003. The Genetic Legacy of the Mongols. 72(3).

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. *The 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-B, Duan Z-Y, Tan P-P, 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-B, Xiao C-J, 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* 5540(10): 428–35.

Web resources

Ensemble https://www.ensembl.org/index.html
ISOGG http://www. Isogg.org
PhyloTree http://www.phylotree.org
Vega http://vega.archive.ensembl.org/index.html

ACKNOWLEDGEMENTS

I'm deeply grateful for all the **donors** who entrusted their samples into our scientific enquires and to all the **anthropologists**, **doctors** and other **colleagues** who went out to the field and collected these samples, and to all our **co-authors** for making the science happen. My sincere gratitude to every one of you, since you make such work possible.

The completion of this thesis took somewhat longer than initially anticipated, throughout the years I have been allowed to learn from many supervisors and colleagues. Thank you all for the wise guidance, for creating excellent possibilities to learn, work, strive and for the superbly good times together.

In the historical appearance: **Toomas**, even though twists of life did not set the work off when planned and our tight collaboration started much later due to other turns of fate, though at distance, you have always been sharp, quick, clear and witty in leading the way. I'm ever grateful for that. **Ene**, thank you very much for picking me up while a bit lost, offering an interesting project to work on, being always there, looking after me and pushing to new projects. **Mait**, when tables turned and new avenues opened, you took the lead. Thank you for the myriad of challenges that had to be faced yesterday or the day before, for the everlasting faith into my capacities despite all signs opposite, for creating the opportunities available to us in the lab and surely for the music. **Siiri**, I'd be lost without you in this thicket of Y-chromosomes, you are always keen to finish things, pushing for the better with the careful eye on details. Thank you and don't worry, it will be alright! **Richard**, always there with encyclopaedic knowledge and wide scope, ever challenging to seek for novelty and setting the scientific standards higher. Thank you for all this and surely for mentioning the "*Goat and Sheep of South Estonia*" as a potential avenue for scientific publishing. **Jüri**, you were my first ever supervisor eons ago, since then you've been a perpetual wake-up caller with the jokes and meters-long lenses, and this, as you know, has been celebrated with loads of yoghurt. **Kristiina**, thank you so much for the constant – *no child will cry, not on my watch* – attitude, for being a role model of a mother and a scientist and always there as a good friend. **Lauri** – music, talks, and the endless lessons of code and server whispering, no sorries, just countless thank-yous!

Throughout the years the numbers of the most lovely people in our lab-family have grown large, its members scattered in many parts of the world. **Gyaneshwer** and **Chandana**, thank you for the endless laughs and bringing your Indian family here. **Phillip** – a paragraph a day, what a marvellous advice on top of many conversations about wonders of life. **Doron**, thank you for bringing in some structure, strategic thinking, and the olive oil. **Lena**, the ant in the house, thank you for the warm nights in your kitchen. **Georgi**, thank you for flying me to the other end of the world once again! To all the members of this wonderful international family of Estonian Biocentre: *Muchas grazias*! *Grazie mille*! Thank you! *Vielen Dank*! *Teşekkür ederim*! Շնորհակալություն! *Muito*

obrigado! আপনাকে ধন্যবাদ! *Puno hvala*! धन्यवाद! Большое спасибо! *Merci beaucoup*! Рəхмəт! *Suur aitähh*! You make the scientific rigour chime together with the warmest laughter.

Erwan, we've been on this galley together, rowing on the bottom deck, but now it's time to change seats, go up on the deck for fresh air and enjoy the view. I'm waiting for you with great anticipation! **Anne-Mai**, you'll be there in a moment!

Sidney and **Jonathan**, the chances are slim for three of us, so different, meeting in that piece of land settled by humans the last. There we were, in Auckland. Sid, thank you very *viel* for your most insightful lessons about the true agonies of arm wrestling with the thesis, outwitted by puns and language yearning. Jo, challenging (the climbing wall) you've taught, with passionate insights to open access movement, math, design and the power of pandas. Thank you both for the trust and the most sincere conversations.

My special deep admiration goes to all the mothers in science. To name a few, not mentioned so far, who have inspired and supported me: **Ann, Jana, Mari-Anne, Helen, Astra, Sirje, Maere, Erika, Anu, Mari, Kärt, Renata, Maris, Elin**. You are amazing!

Countless hours are spent in the lab, but the wonderful people outside keep reminding about the life beyond. *Punase Kontsakinga Ordu*, the Order of the Red Stiletto, has become a true source of daily joy. **Merje** and the small engineers, **Kati**, **Aleks**, **Reti**, and **Lennart** – once neighbours, always neighbours. You organise all the trips and hikes and spontaneous birthday parties and visits at random times. **Tuuli**, our friendship has for long grown out from the lab walls into this small support community. **Merike** and **Aire**, the talks and the walks, essence of life! **Leho**, my bro from different parents. **Pille**, **Siim**, **Tõnis**, **Siki**, **Mari**, **Ingrid**, **Kristiina**, **Kadri** – you have been there from the very start of our endeavour. *Even if the outcome of my university years would have been meeting you guys, I'd be totally satisfied* (Värv 2011). (I'm almost living up to Signe with the length of the acknowledgements! And, is the next Christmas party in June or July?). **Herne Street People**, what would life be without your garden and kitchen where the heated debates about life, science and starfish take place? **Hedi**, **Renee**, **Marrit** – some people just are part of the constellation. **Godi**, the years-long in depth discussions on the comparative study of Norwegian and Estonian landscapes have filled many hours with enormous warmth. Thank you all for being there!

It is hard to find words to fully express the gratitude to my closest family. I am thankful to my parents, **Mati** and **Lemmi**, who have been there for us, taught me the respect for history, showed how work brings gratification and talent responsibilities. **Maarja** and **Miki**, you are the siblings I need. With deep warmth I marvel the image of my late granny **Salme**, for always being there, full of love and no expectations, and for intertwining the beauty of Estonian nature with that of our language. **Martin**, you've gently been there. Always. Taking care of everything so naturally, your love and trust have nourished our life. Thnak yuo for a patients :P. **Kusti**, **Jaagup** and **Anni** – what to say, but to

loosely quote: science can be fun, but it is nothing compared to a smile on a child's face (Metspalu 2005), nor to hugging and looking into the eyes of children. When realizing at some point that these eyes are gazing down at you, the awe grows. Our marvellous children, you have brought so much joy, fun and striving for better into my life. Thank you for coming, being here and becoming my main teachers! Surely, **Täpi**, from your eyes I know you never had a meal, but the wagging tail tells it is the utmost time for another walk. Time to go and say thanks to all I forgot to mention.

References:

Metspalu M, Through the course of prehostory in India: tracing the mtDNA trail. 2005. *Dissertationes Biologicae Universitatis Tartuensis*. Tartu University Press.

Värv S, Studies on the mechanisms of RNA polymerase II-dependent transcription elongation. 2011. *Dissertationes Biologicae Universitatis Tartuensis*. Tartu University Press.

PUBLICATIONS

CURRICULUM VITAE

Name: Monika Karmin
Date of Birth: January 23, 1979
Family: in partnership, three children
Contacts: University of Tartu, Institute of Genomics, Riia 23b, 51010 Tartu, Estonia, +372 737 5050
E-mail: monika.karmin@ut.ee

Education:

2005–2014	Doctoral studies, University of Tartu, Faculty of Science and Technology, Institute of Molecular and Cell Biology, Chair of Evolutionary Biology
2002–2005	University of Tartu, Faculty of Biology and Geography, Institute of Molecular and Cell Biology, Chair of Evolutionary Biology. "Human Mitochondiral DNA Haplogroup R in India: Dissecting the Phylogenetic Tree of South Asian-Specific Lineages", Supervisor Ene Metspalu. *MSc*
1998–2003	University of Tartu, Faculty of Biology and Geography, Institute of Molecular and Cell Biology, Chair of Evolutionary Biology, *BSc*

Professional Employment:

2018–...	University of Tartu, Institute of Genomics, Research Fellow of Population Genetics
2018–2018	Massey University, Palmerston North, New Zealand, Visiting Researcher
2016–2016	University of Auckland, New Zealand, Visiting Researcher
2016–2017	Estonian Biocentre, Researcher
2009–2015	Estonian Biocentre, Extraordinary Researcher
2007–2007	Quattromed HTI Laborid OÜ, lab assistant

Distinctions and Awards

2018	Massey University International Visitor Research Fund grantee
2017	Estonian State science prize, member of the group of laureates
2016	NEFREX grantee as a visiting scientist to University of Auckland, New Zealand
2010	EBC Women in Science grantee

Creative work and additional career information:

2015	Collaboration in renewing of the section on Evolution, molecular and cellular biology in the main exhibit of the Natural History Museum of University of Tartu.
2012–2013	Supervising biology workshops in a project "Workshops of physics, chemistry and biology"

2010–… Lecturer and instructor of practical work on the bachelor and masters level in the University of Tartu. Topics: Immunology and Immunogenetics, Evolutionary Biology, Population Genetics.

Attended courses:

2015 Computational Molecular Evolution; Wellcome Trust Advance Courses in Sanger Center, UK.
2015 Illumina NextSeq 500 in-house introduction course; Tartu.
2012 Statistical Practice in Epidemiology using R; Pärnu, Estonia.
2011 Immunoregulation: Cytokines and interferons; Tallinn Technical University.
2011 Exome Sequencing, Genotyping and ArrayCGH: Technologies for Genome Analysis; Wellcome Trust Advance Courses in Sanger Center, UK.
2010 Course of Advance Sequencing Technologies and Applications; Cold Spring Harbor Laboratory, USA.

Publications:

Tambets K, Yunusbayev B, Hudjashov G, Ilumäe AM, Rootsi S, Honkola T, Vesakoski O, Atkinson Q, Skoglund P, Kushniarevich A, Litvinov S, Reidla M, Metspalu E, Saag L, Rantanen T, **Karmin M**, Parik J, Zhadanov SI, Gubina M, Damba LD, Bermisheva M, Reisberg T, Dibirova K, Evseeva I, Nelis M, Klovins J, Metspalu A, Esko T, Balanovsky O, Balanovska E, Khusnutdinova EK, Osipova LP, Voevoda M, Villems R, Kivisild T, Metspalu M. 2018. **Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations.** *Genome Biology*. Sep 21;19(1):139.

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 JR, 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,* Jan 29; 8(1):1823.

Behar DM, Saag L, **Karmin M**, Gover MG, Wexler JD, Fernanda L, Greenspan E, Kushniarevich A, Davydenko O, Sahakyan H, Yepiskoposyan L, Boattini A, Sarno S, Pagani L, Carmi S, Tzur S, Metspalu E, Bormans C, Skorecki K, Metspalu M, Rootsi S, Villems R. 2017. **The genetic variation in the R1a clade among the Ashkenazi Levites' Y chromosome.** *Scientific Reports*, Nov 2;7(1):14969.

Chaubey G, Ayub Q, Rai N, Prakash S, Mushrif-Tripathy V, Mezzavilla M, Pathak AK, Tamang R, Firasat S, Reidla M, **Karmin M**, Rani DS, Reddy AG, Parik J, Metspalu E, Rootsi S, Dalal K, Khaliq S, Mehdi SQ, Singh L, Metspalu M, Kivisild T, Tyler-Smith C, Villems R, Thangaraj K. 2017. "**Like sugar in milk": reconstructing the genetic history of the Parsi population.** *Genome Biology*, Jun 14;18(1):110.

Sahakyan H, Hooshiar KB, Tamang R, Kushniarevich A, Francis A, Costa MD, Pathak AK, Khachatryan Z, Sharma I, vanOven M, Parik J, Hovhannisyan H, Metspalu E, Pennarun E, **Karmin M**, Tamm E, Tambets K, Bahmanimehr A, Reisberg T, Reidla M, Achilli A, Olivieri A, Gandini F, Perego UA, Al-Zahery N, Houshmand M, Sanati MH, Soares P, Rai E, Šarac J, Šarić T, Sharma V, Pereira L, Fernandes V, Černý V, Farjadian S, Singh DP, Azakli H, Üstek D, Ekomasova (Trofimova) N, Kutuev I, Litvinov S, Bermisheva M, Khusnutdinova EK, Rai N, Singh M, Singh VK, Reddy AG, Tolk H, Cvjetan S, Lauc LB, Rudan P, Michalodimitrakis EN, Anagnou NP, Pappa KI, Golubenko MV, Orekhov V, Borinskaya SA, Kaldma K, Schauer MA, Simionescu M, Gusar V, Grechanina E, Govindaraj P, Voevoda M, Damba L, Sharma S, Singh L, Semino O, Behar DM, Yepiskoposyan L, Richards MB, Metspalu M, Kivisild T, Thangaraj K, Endicott P, Chaubey G, Torroni A, Villems R. 2017. **Origin and spread of human mitochondrial DNA haplogroup U7**. *Scientific Reports*, Apr 7;7:46044.

Pagani L, Lawson DJ, Jagoda E, Mörseburg A, Eriksson A, Mitt M, Clemente F, Hudjashov G, DeGiorgio M, Saag L, Wall JD, Cardona A, Mägi R, Wilson Sayres MA, Kaewert S, Inchley C, Scheib CL, Järve M, **Karmin M**, Jacobs GS, Antao T, Iliescu FM, Kushniarevich A, Ayub Q, Tyler-Smith C, Xue Y, Yunusbayev B, Tambets K, Mallick CB, Saag L, Pocheshkhova E, Andriadze G, Muller C, Westaway MC, Lambert DM, Zoraqi G, Turdikulova S, Dalimova D, Sabitov Z, Sultana GNN, Lachance J, Tishkoff S, Momynaliev K, Isakova J, Damba LD, Gubina M, Nymadawa P, Evseeva I, Atramentova L, Utevska O, Ricaut F, Brucato N, Sudoyo H, Letellier T, Cox MP, Barashkov NA, Škaro V, Mulahasanovic´ L, Primorac D, Sahakyan H, Mormina M, Eichstaedt CA, Lichman DV, Abdullah S, Chaubey G, Wee JT, Mihailov E, Karunas A, Litvinov S, Khusainova R, Ekomasova N, Akhmetova V, Khidiyatova I, Marjanović D, Yepiskoposyan L, Behar DM, Balanovska E, Metspalu A, Derenko M, Malyarchuk B, Voevoda M, Fedorova SA, Osipova LP, Lahr MM, Gerbault P, Leavesley M, Migliano AB, Petraglia M, Balanovsky O, Khusnutdinova EK, Metspalu E, Thomas MG, Manica A, Nielsen R, Villems R, Willerslev E, Kivisild T, Metspalu M. 2016. **Genomic analyses inform on migration events during the peopling of Eurasia**. *Nature*, 538 (7624), 238−242.10.1038/nature19792.

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, Balanovska E, Behar DM, Balanovsky O, Kivisild T, Underhill PA, Villems R, Rootsi S. 2016. **Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families.** *The American Journal of Human Genetics* 99(1):163–73.

Karmin M, Saag L, Vicente M, Wilson Sayres MA, Järve M, Gerst Talas U, Rootsi S, Ilumäe A-M, Mägi R, Mitt M, Pagani L, Puurand T, Faltyskova Z, Clemente F, Cardona A, Metspalu E, Sahakyan H, Yunusbayev B, Hudjashov G, DeGiorgio M, Loogväli E-L, Eichstaedt C, Eelmets M, Chaubey G, Tambets K, Litvinov S, Mormina M, Xue Y, Ayub Q, Zoraqi G, Korneliussen TS, Akhatova F, Lachance J, Tishkoff S, Momynaliev K, Ricaut F, Kusuma P, Razafindrazaka H, Pierron D, Cox MP, Sultana GNN, Willerslev R, Muller C, Westaway M, Lambert D, Skaro V, Kovačević L, Turdikulova S, Dalimova D, Khusainova R, Trofimova N, Akhmetova V, Khidiyatova I, Lichman DV, Isakova J, Pocheshkhova E, Sabitov Z, Barashkov NA, Nymadawa P, Mihailov E, Seng JWT, Evseeva I, Migliano AB, Abdullah S, Andriadze G, Primorac D, Atramentova L, Utevska O, Yepiskoposyan L, Marjanović D, Kushniarevich A, Behar DM, Gilissen C, Vissers L, Veltman JA, Balanovska E, Derenko M, Malyarchuk B, Metspalu A, Fedorova S, Eriksson A, Manica A, Mendez FL, Karafet TM, Veeramah KR, Bradman N, Hammer MF, Osipova LP, Balanovsky O, Khusnutdinova EK, Johnsen K, Remm M, Thomas MG, Tyler-Smith C, Underhill PA, Willerslev E, Nielsen R, Metspalu M, Villems R, Kivisild T. 2015**. A recent bottleneck of Y chromosome diversity coincides with a global change in culture.** *Genome Research* 25(4):459–466.

Chaubey G, Singh M, Crivellaro F, Tamang R, Nandan A, Singh K, Sharma VK, Pathak AK, Shah AM, Sharma V, Singh VK, Selvi Rani D, Rai N, Kushniarevich A, Ilumäe AM, **Karmin M**, Phillip A, Verma A, Prank E, Singh VK, Li B, Govindaraj P, Chaubey AK, Dubey PK, Reddy AG, Premkumar K, Vishnupriya S, Pande V, Parik J, Rootsi S, Endicott P, Metspalu M, Lahr MM, van Driem G, Villems R, Kivisild T, Singh L, Thangaraj K. 2014. **Unravelling the distinct strains of Tharu ancestry.** *European Journal of Human Genetics*, 22(12), 1404–1412.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW Jr, Rasmussen S, Moltke I, Albrechtsen A, Doyle SM, Poznik GD, Gudmundsdottir V, Yadav R, Malaspinas A-S, Stockton White V S, Allentoft ME, Cornejo OE, Tambets K, Eriksson A, Heintzman PD, **Karmin M**, Korneliussen TS, Meltzer DJ, Pierre TL, Stenderup J, Saag L, Warmuth VM, Lopes MC, Malhi RS, Brunak S, Sicheritz-Ponten T, Barnes I, Collins M, Orlando L, Balloux F, Manica A, Gupta R, Metspalu M, Bustamante CD, Jakobsson M, Nielsen R, Willerslev E. 2014. **The genome of a Late Pleistocene human from a Clovis burial site in western Montana.** *Nature*, 506(7487), 225–29.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW Jr, Orlando L, Metspalu E, **Karmin M**, Tambets K, Rootsi S, Mägi R, Campos PF, Balanovska E, Balanovsky O, Khusnutdinova E, Litvinov S, Osipova LP, Fedorova SA, Voevoda MI, DeGiorgio M, Sicheritz-Ponten T, Brunak S, Demeshchenko S, Kivisild T, Villems R, Nielsen R, Jakobsson M, Willerslev E. 2014. **Upper Palaeolithic Siberian**

genome reveals dual ancestry of Native Americans. *Nature*, 505(7481), 87–91.

Rai N, Chaubey G, Tamang R, Pathak AK, Singh VK, **Karmin M**, Singh M, Rani DS, Anugula S, Yadav BK, Singh A, Srinivasagan R, Yadav A, Kahsyap M, Narvariya S, Reddy AG, VanDriem G, Underhill PA, Villems R, Kivisild T, Singh L, Thangaraj K. 2012. **The Phylogeography of Y-Chromosome Haplogroup H1a1a-M82 Reveals the Likely Indian Origin of the European Romani Populations.** *PLoS ONE*, 7, e48477.

Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, Reidla M, Tamm E, **Karmin M**, Reisberg T, Kashani B, Perego U, Carossa V, Gandini F, Pereira J, Soares P, Angerhofer N, Rychkov S, Al-Zahery Nadia, Carelli V, Sanati M, Houshmand M, Hatina J, Macaulay V, Pereira L, Woodward S, Davies W, Gamble C, Baird D, Semino O, Villems R, Torroni A, Richards M. 2012. **Mitochondrial DNA signals of Late Glacial re-colonisation of Europe from Near Eastern refugia.** *American Journal of Human Genetics*, 90(5), 915–24.

Chaubey G, Metspalu M, Choi Y, Mägi R, Gallego RI, Soares P, van Oven M, Behar DM, Rootsi S, Hudjashov G, Basu MC, **Karmin M**, Nelis M, Parik J, Goverdhana RA, Metspalu E, vanDriem G, Xue Y, Tyler-Smith C, Thangaraj K, Singh L, Remm M, Richards MB, Mirazon LM, Kayser M, Villems R, Kivisild T. 2011. **Population Genetic Structure in Indian Austroasiatic speakers: The Role of Landscape Barriers and Sex-specific Admixture.** *Molecular Biology and Evolution*, 28(2), 1013–24.

Rasmussen M, Guo X, Wang Y, Lohmueller K E, Rasmussen S, Albrechtsen A, Skotte L, Lindgreen S, Metspalu M, Jombart T, Kivisild T, Zhai W, Erikson A, Manica A, Orlando L, De La Vega F, Tridico S, Metspalu E, Nielsen K, Avila-Arcos MC, Moreno-Mayar J V, Muller C, Dortch J, Gilbert M T P, Lund O, Wesolowska A, **Karmin M**, Weinert L A, Wang B, Li J, Tai S, Xiao F, Hanihara T, Driem G, Jha A R, Ricaut F-X, Knijff P, Migliano A B, Gallego-Romero I, Kristiansen K, Lambert D M, Brunak S, Forster P, Brinkman B, Nehlich O, Bunce M, Richards M, Gupta R, Bustamante C D, Krogh A, Foley R A, Lahr MM, Balloux F, Sicheritz-Ponten T, Villems R, Nielsen R, Jun W, Willerslev E. 2011. **An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia**. *Science*, 334(6052), 94–8.

Chaubey G, **Karmin M**, Metspalu E, Metspalu M, Selvi-Rani D, Singh VK, Parik J, Solnik A, Naidu BP, Kumar A, Adarsh N, Mallick CB Trivedi B, Prakash S, Reddy R, Shukla P, Bhagat S, Verma S, Vasnik S, Khan I, Barwa A, Sahoo D, Sharma A, Rashid M, Chandra V, Reddy AG, Torroni A, Foley RA, Thangaraj K, Singh L, Kivisild T, Villems R. 2008. **Phylogeography of mtDNA haplogroup R7 in the Indian peninsula.** *BMC Evolutionary Biology*, 8:227.

Chaubey G, Metspalu M, **Karmin M**, Thangaraj K, Rootsi S, Parik J, Solnik A, Rani D,S, Singh V,K, Naidu B,P, Reddy A,G, Metspalu E, Singh L, Kivisild T, Villems R. 2008**. Language Shift by Indigenous Population: A Model**

Genetic Study in South Asia. *International Journal of Human Genetics*, 8, 41–50.

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, **Karmin M**, Behar DM, Gilbert MTP, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R. 2004. **Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans.** *BMC Genetics,* 5, 26.

ELULOOKIRJELDUS

Nimi: Monika Karmin
Sünniaeg: 23. jaanuar 1979
Perekond: elukaaslane, kolm last
Kontakt: Genoomika Instituut, Tartu Ülikool, Riia 23, Tartu 51010, +372 737 5050
E-post monika.karmin@ut.ee

Haridus:

2005–2014	Doktoriõpinguid. Tartu Ülikool, Loodus- ja tehnoloogiateaduskond, Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituut, Evolutsioonilise bioloogia õppetool
2002–2005	Tartu Ülikool, Bioloogia-geograafiateaduskond, Molekulaar- ja rakubioloogia instituut, *MSc,* (juh) Ene Metspalu, "Human Mitochondiral DNA Haplogroup R in India: Dissecting the Phylogenetic Tree of South Asian-Specific Lineages"
1998–2003	Tartu Ülikool, Bioloogia-geograafiateaduskond, Molekulaar- ja rakubioloogia instituut, *BSc*

Töökohad ja ametid:

2018–...	Tartu Ülikool, Tartu Ülikooli genoomika instituut, populatsioonigeneetika teadur
2018–2018	Massey Ülikool, Palmerston North, Uus-Meremaa, külalisteadur
2016–2016	Aucklandi Ülikool, Auckland, Uus-Meremaa, külalisteadur
2016–2017	Eesti Biokeskus, teadur
2009–2015	Eesti Biokeskus, erakorraline teadur
2007–2007	Quattromed HTI Laborid OÜ, laborant

Tunnustused ja auhinnad:

2018	Massey Ülikooli rahvusvahelise külalisteaduri stipendium
2017	Riigi teaduspreemia, laureaatide rühma liige
2016	NEFREX stipendium Aucklandi Ülikooli külastamiseks
2010	EBK Naised teaduses stipendium

Teenistuskäigu lisainfo:

2015	Kaastöö "Evolutsioon, molekulaar- ja rakubioloogia" sektsiooni uuendamisele TÜ Loodusmuuseumis.
2012–2013	Bioloogia õpikodade läbiviimine projektis "Füüsika, keemia ja bioloogia õpikojad"
2010–...	Loengud ja praktikumide juhendamised Tartu Ülikooli bakalaureuse ja magistritaseme kursustel: immunoloogia ja immunogeneetika, evolutsiooniline bioloogia, populatsioonigeneetika

Kursused:

2015 Arvutuslik molekulaarne evolutsioon; Wellcome Trust Advance Courses in Sanger Centre, Ühendkuningriik.
2015 Illumina NextSeq 500 tutvustav kursus; Tartu.
2012 Statistika R-iga epidemioloogias; Pärnu, Eesti.
2011 Immunoregulatsoon: tsütokiinid ja interferoonid, TTÜ.
2011 Eksoomisekveneerimine, genotüpiseerimine ja kiibiCGH: genoomianalüüsi tehnoloogiad; Wellcome Trust Advance Courses in Sanger Centre, Ühendkuningriik.
2010 Kaasaegsed sekveneerimistehnoloogiad ja nende rakendused; Cold Spring Harbor Laboratory, Ameerika Ühendriigid.

Teaduspublikatsioonid:

loetletud ingliskeelse 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. **Kaasik, Krista**. 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 Usküla**. 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 boreonemoral 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.