DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
361

ERWAN PENNARUN

Meandering along the mtDNA phylogeny; causerie and digression about what it can tell us about human migrations

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
361

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
361

ERWAN PENNARUN

Meandering along the mtDNA phylogeny; causerie and digression about what it can tell us about human migrations

UNIVERSITY OF TARTU
Press

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

Dissertation was accepted for the commencement of the degree of Doctor of Philosophy in Molecular and Cellular Biology on 30th of August 2019 by the Council of Institute of Molecular and Cell Biology, Faculty of Science and Technology, University of Tartu.

Supervisors: Professor Richard Villems
Institute of Molecular and Cell Biology
Department of Evolutionary Biology
Institute of Genomics
University of Tartu

Mait Metspalu, PhD,
Institute of Genomics
University of Tartu

Opponent: Assistant Professor Rosa Irene Fregel Lorenzo
Department of Biochemistry, Microbiology,
Cellular Biology and Genetics
Universidad de La Laguna

Commencement: Room No 105, 23B Riia St, Tartu, on 7th of October 2019 at 14:15.

ISSN 1024-6479
ISBN 978-9949-03-164-1 (print)
ISBN 978-9949-03-165-8 (pdf)

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

LIST OF ORIGINAL PUBLICATIONS

The current dissertation is based on the following publications, referred therein by their Roman numerals:

I. Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, **Pennarun E**, Parik J, Geberhiwot T, Usanga E, Villems R. (2004). Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet. Nov;75(5):752–70.

II. Richard C*, **Pennarun E***, Kivisild T, Tambets K, Tolk HV, Metspalu E, Reidla M, Chevalier S, Giraudet S, Lauc LB, Pericić M, Rudan P, Claustres M, Journel H, Dorval I, Müller C, Villems R, Chaventré A, Moisan JP. (2007). An mtDNA perspective of French genetic variation. Ann Hum Biol. Jan-Feb;34(1):68–79.
* Authors equally contributed to the study.

III. **Pennarun E**, Kivisild T, Metspalu E, Metspalu M, Reisberg T, Moisan JP, Behar DM, Jones SC, Villems R. (2012). Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa. BMC Evol Biol. 2012 Dec 3;12:234.

IV. Di Cristofaro J*, **Pennarun E***, Mazières S, Myres NM, Lin AA, Temori SA, Metspalu M, Metspalu E, Witzel M, King RJ, Underhill PA, Villems R, Chiaroni J. (2013). Afghan Hindu Kush: where Eurasian sub-continent gene flows converge. PLoS One. Oct 18;8(10)
* Authors equally contributed to the study

My contribution to these above-mentioned articles is as follow:

Ref. I. a) performed some of the statistical analyses b) participated in the preparation of the manuscript;

Ref. II. a) performed 1/3rd of the sequencing and genotyping b) performed the analysis of the data with C. Richard c) wrote the manuscript with C. Richard;

Ref. III. a) had a key role in the design of the study b) did all the sequencing and genotyping c) performed all the analyses d) wrote the manuscript with T. Kivisild, S. C. Jones and R. Villems;

Ref. IV. a) performed all the mtDNA sequencing and genotyping b) did most of the analyses for mtDNA and all the analyses for autosomal data c) helped drafting the manuscript d) wrote more specifically the parts about the mtDNA and autosomal aspects of the study.

ABBREVIATIONS

AMH	Anatomically Modern Human(s)
bp/kbp	Base Pair/thousand (Kilo) Base Pairs
BSP	Bayesian Skyline Plot
hg(s)	Haplogroup(s)
HVS(R)-I/HVS(R)-II	$1^{st}/2^{nd}$ HyperVariable Segment(Region)
KYA, Myr	thousand (Kilo-) Years Ago, Million Years
(T)MRCA	(Time to the) Most Recent Common Ancestor
mtDNA	mitochondrial DeoxyriboNucleic Acid
N_e	Effective population size
np(s)	Nucleotide Position(s)
NGS	Next Generation Sequencing
PCR	Polymerase Chain Reaction
RFLP	Restriction Length Polymorphism
SNP	Single Nucleotide Polymorphism
OoA	"Out of Africa"
chrY	Y chromosome

Definitions of some terms used in this dissertation

Haplotype	Combination of several mtDNA polymorphisms in a defined allelic state
Haplogroup	In the context of mtDNA and Y-Chromosome phylogenetic studies, the monophyletic cluster (clade) of haplotypes sharing an array of defining polymorphism(s)
Coalescence time	Time estimate to MRCA

1. INTRODUCTION

(Some) Humans tend to be of very curious nature. Although that trait is not particular to Humans, one can say that the scopes of the questions in which our curiosity delves are manifold, from very concrete, down to earth, to downright abstract ones. From Where is my food? Is my pension fund performing well? for the latest western edition of the Anatomically Modern Humans (AMH), to Where do we come from? It is likely that the latter question is a rather ancient one. The tools at hands to answer it are various, from peremptory claims based on mere whimsical constructions of the mind to testing such ones with a more scientific approach, based on observations, experiments, hypotheses testing, modelling and the likes. From the biological perspective, for a long period, only the phenotypical diversity of humans was available to scrutiny in order to address it. The progress in the understanding of how the cell is working, of how heredity is stored and passed along; those were major advancements that saw light only relatively recently. One cannot avoid giving credit to the work of James D. Watson and Francis Crick, based on those of the oft overlooked Rosalind Franklin and Raymond Gosling. Their discovery of the DNA structure and its putative replicative nature paved the way for many achievements in molecular biology in the subsequent years, as these finally provided the opportunity for studying the genotype rather than the phenotype. Further, a definite turning point helping in a readily access to the genotype was the implementation of the Polymerase Chain Reaction by Kary Mullis around the mid-80's, following on the footsteps of the technique of Sanger sequencing. Combined together, it freed researchers from the need of bleeding subjects to get sufficient DNA for their studies; an improvement that surely was and is of great comfort to many a participant of scientific and medical studies. Another technological development that helped quenching this thirst for blood is Next Generation Sequencing, the massively parallel sequencing of a multitude of tagged fragments giving access to virtually the whole genome. These are but technical achievements of profound implications, and along the way, theories equally important were formulated.

One particularly important and relevant theory to the field of population genetics is the neutral theory of molecular evolution introduced by Motoo Kimura that holds that, at the molecular level, most of the variation found between and within species is caused by genetic drift and not natural selection. This assertion caused quite a stir amongst the scientific community but it nonetheless provides a convenient null-hypothesis. Less controversial is the coalescent theory, which, by modelling how gene variants are sampled backward in time, provides also a framework for testing different hypothesis of demographic parameters.

Armed with this knowledge and these technical advancements, we can tackle that question of origin in a more befitting manner, and also broaden our inquiries to more specific questions. If nowadays accessing the full genome,

and hence almost the entire genetic information embedded in it, is readily doable and will soon likely becoming standard thanks to cost reductions and technology developments, it is only the genetic aspect of the question. The genetic variation of the extant populations has been generated by diverse and opposite forces. It is therefore instructive to resort to other disciplines to shed some lights on various events that could have influenced shaping the present day genetic diversity. To supplement and strengthen the genetic findings, to borrow and flip around the words of the late Luigi Cavalli-Sforza (Cavalli-Sforza *et al.* 1991), one often calls on archaeology, history and linguistics, as well as on paleoclimatology. Whilst providing non-biological context, they are also valuable in further refining the questions that can be explored. This synergistic combination of genetics and multidisciplinary inputs habitually gives rise to a more comprehensive analysis of the results.

This dissertation presents four such inclusive studies (or narratives, pending artistic license is permitted) using genetic data, all of them extensively relying on maternally inherited mitochondrial DNA. Ref. I will bring us to Ethiopia in East Africa and across the Bab-el-Mandeb into Yemen. After describing the gene pool of these two countries using high-phylogenetic resolution, we set out to track the recent flow that has occurred around this region. Ref. II brings us to Europe and uses a similar approach to increase the French mtDNA dataset and to place its variation into the broader European context. We also explore further potential connections that may be revealed by the haplogroups and specific haplotypes found in Brittany and the Basque country. With Ref. IV, we go to Central Asia and more specifically in the Afghan Hindu Kush. Once more, we describe in details the mtDNA diversity present there, as well as that of the Y chromosome. We also report a survey of over 600,000 single nucleotide polymorphisms found on the autosomes. Combining these data, we address the question of whether Central Asia has been or not a place from which human migrations radiated. We finally go back to Africa and around the Mediterranean Basin in Ref. III. In this study, we focus our attention on two mitochondrial haplogroups, M1 and U6. By characterising fully their phylogenies and further genotyping a broader set of samples, we set to revisit the proposed concomitance of their temporal and geographical dispersion. We are also able to formally test the hypothesis which posits that the dispersal of these haplogroups is correlated with that of Afro-Asiatic languages.

It is hoped that this thesis and the papers constituting it will manage to give the readers some insights into the potential and limitations of using mtDNA and its phylogeny within the framework of population genetics.

2. LITERATURE OVERVIEW

Before bringing our attention to the studies forming the basis of the present dissertation, I shall present and briefly discuss below few essential biological characteristics of the different genomes used in them, followed by the introduction of certain of the tools and methods typically employed in population genetics studies.

2.1. Of haploid and diploid parts of the human genome

The genome that constitutes the complete genetic information of humans comprises two parts: the nuclear genome and the mitochondrial genome. The nuclear genome is composed of twenty-three pairs of chromosomes found, as its name implies, within the nucleus of cells. Chromosomes 1 to 22 are also known as the autosomes, the twenty-third pair being composed by the XY sex chromosomes. The mitochondrial genome is located in the organelle found in virtually every cell in the human body that is the mitochondrion. Whilst both genomes are composed of the same basic building blocks forming the DNA, they differ in some aspects that are crucial for population genetics studies.

2.1.1. The female line archives of the mitochondria

The mitochondrial genome is a compact double stranded circular molecule of ≈16,659 bp first sequenced entirely in 1981 (Anderson *et al.* 1981). It is usually referred to as the Cambridge Reference Sequence (CRS), sequence later revised, with 11 nucleotides corrected (Andrews *et al.* 1999) and known as the revised CRS (rCRS). There is also the Reconstructed Sapiens Reference Sequence (RSRS) introduced in 2012 (Behar *et al.* 2012b), which aims at solving some of the issues regarding the root of the human mtDNA tree inherent to the use of the rCRS (see section 2.2.1.1 below). The mtDNA genome itself contains 37 genes, with no introns in them and almost no intergenic non-coding nucleotides, except for the 1.1 kb of the displacement loop. The displacement loop is involved in the regulation of the transcription and replication of the mitochondrial genome (Chinnery 2006) and it contains the segments called HVS-I, HVS-II and HVS-III, also referred to as HVR in some other studies. These are particularly interesting in the context of evolutionary studies as their mutation rate is even higher than that of the rest of the molecule (see section 2.2.2). Regarding the genes, they encode 22 transfer RNAs, 2 ribosomal RNAs (12s and 16s) and 13 proteins of the electron transport chain; 7 of complex I or NADH dehydrogenase (ND1-ND6), the cytochrome *b* of complex III, 3 of complex IV or cytochrome *c* oxidase (COX1-COX3) and 2 of complex V or mitochondrial ATPase (ATP6 and ATP8).

Several mtDNA molecules (≈6–10) are packed into stable protein-DNA macrocomplexes called nucleoids, primarily associated with the inner mitochondrial membrane (Holt *et al.* 2007; Wang and Bogenhagen 2006). Although

the mitochondrion possesses some of the main mechanisms to repair DNA damages, like the Base Excision Repair (BER) (Pinz and Bogenhagen 1998; Szczesny *et al.* 2008), the Single-Strand Break Repair (SSB) (Rossi *et al.* 2009) or the Mismatch Repair (MMR) (Mason *et al.* 2003), these systems are likely overwhelmed and not sufficient to counteract the damages caused by the oxidative stress connected to the Respiratory Chain complex and the Reactive Oxygen Species (ROS) it engenders (Tuppen *et al.* 2010). Combined with a higher turnover rate, its specific way of replication whereby the DNA molecule spends a longer time in a more vulnerable single strand, this leads to a faster mutation rate of the mitochondrial genome, over an order of magnitude as compared to that of the autosomes, from 1.9×10^{-8} for 3rd position of the codon for protein coding and to 8.2×10^{-9} for RNA genes in mtDNA (Soares *et al.* 2009) to 0.5×10^{-9} on average for the autosomal genome (Scally and Durbin 2012). And within the mitochondrial genome, the HVS-I and HVS-II regions themselves have a mutation rate more than ten times that of the others mitochondrial compartments, with 1.6×10^{-7} and 2.3×10^{-7} respectively (Soares et al. 2009).

One other very important feature of the mtDNA for population genetics studies beside those fast mutation rates is that it is inherited only through the mother, from a population of molecules present in the oocyte prior to fertilisation (Giles *et al.* 1980). There have been however studies that brought to the fore the dreaded potential paternal inheritance/leakage (Luo *et al.* 2018; Schwartz and Vissing 2002), but their validity has been questioned because replication was not possible (Filosto *et al.* 2003; Schwartz and Vissing 2004; Taylor *et al.* 2003) or some experimental issues have been raised [(Lutz-Bonengel and Parson 2019), response in (Luo *et al.* 2019)]. Besides these marginal instances, it was shown that any paternal mtDNA found in the zygote is very quickly eliminated, thus excluding the potential occurrence of recombination between the parental mtDNAs (see (Pyle *et al.* 2015) and the accompanying perspective by (Carelli 2015)). This lack of recombination is shared with the Y chromosome (albeit due to a different mechanism, see section 2.1.2 below), in contrasting aspect to the other autosomes, and this absence of recombination gives access, in essence, to the unshuffled female and male genetic lines, as inserts A and B on Figure 1 exemplify. Attempting to reconstruct the genealogy of each chromatid (an Ancestral Recombination Graph) is rapidly intractable as the number of chromatids and/or their length increase. There exist nonetheless some methods which have raised to the challenge (notably (Rasmussen *et al.* 2014)).

Equally relevant is the case of heteroplasmy. It is a state where distinct mtDNA haplotypes cohabit within a mitochondrion (or an organism). Different methods exist to detect this state, each with different detection threshold and caveats (see (Duan, Tu and Lu 2018) for a review and (Li *et al.* 2010) as an example addressing the use of Next Generation Sequencing). The presence of different mtDNA molecules can cause diseases when a disease-causing variant exceeds a certain ratio compare to the wild-type (Tuppen et al. 2010; Wallace and Chalkia 2013), but it can also create problem with the phylogenetic

reconstruction. The haplotypes are frequently closely related, hinting that the variants came about by mutation along the same female lineage (Avise 2000). An extensive and import study of human heteroplasmies has been recently published (Wei *et al.* 2019), showing that selection acts on the germline. For a variant to be passed on, it needs to be present within the germline. From the oogonia to the primary oocytes, first stage of the female gametogenesis, the mtDNA genome undergoes a bottleneck, leading to an increase in the intra-cellular variance, that is heteroplasmies are more pronounced. The study shows that selection acts on heteroplasmies and that, moreover, the actual genetic background of the nuclear genome has on influence on them.

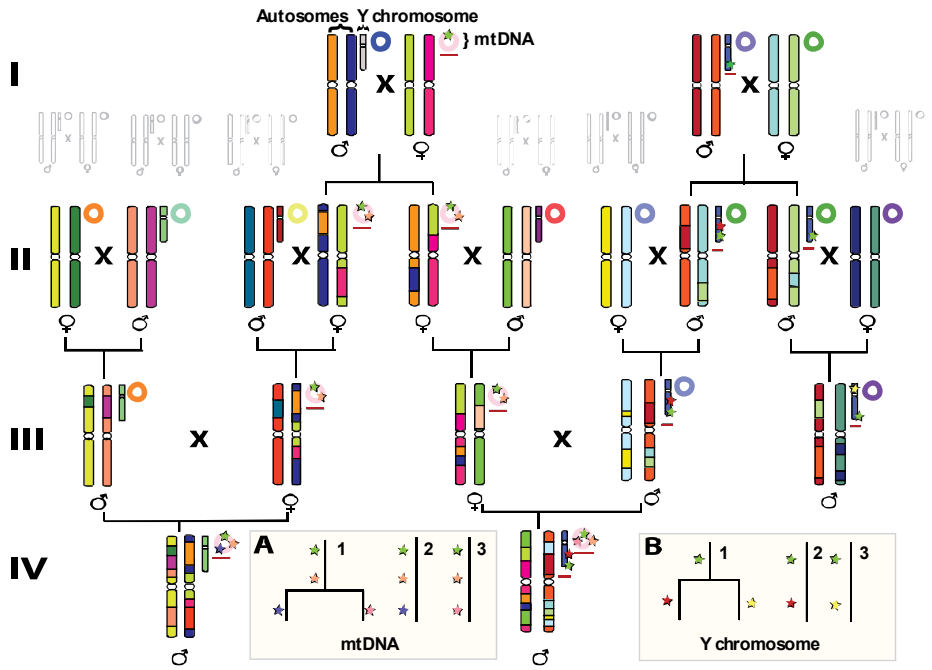

Figure 1. Schema of the haploid inheritance of mtDNA and Y chromosome. mtDNA and the Y chromosome accumulate mutations sequentially (indicated by little stars) through the generations (roman numerals on the left) and do not undergo recombination between each generation contrary to the autosomes where crossing over during meiosis shuffles the chromatids (inclusions of the coloured bits of the sister chromatid). Insert A shows the several reconstructed phylogenies of the mtDNA molecule(s) underlined. For instance, if both females in generation IV are sampled, a tree like the one numbered 1 could be reconstructed. But if we were to sample only the female on the left of insert A, we would obtain the tree 2 (and any order of the mutations would be equally parsimonious) and tree 3 if only the female to the right of the insert were sampled. Likewise, insert B shows the several reconstructed phylogenies of the Y chromosome that would be obtain if sampling the males in generation III whose Y chromosome is underlined; the phylogenetic tree 1 if both are sampled, or 2 or 3 if only the male on the left or on the right are respectively sampled. Note that the above remark about the order of the mutations applies here as well.

Another feature that mtDNA shares with Y chromosome is its effective population size (N_e). Briefly, N_e is a concept that was developed in the 1930s by Sewall Wright (concept reviewed in (Charlesworth 2009) for instance) in order to calculate the amount of genetic drift experienced by a population. N_e is equal to the size of an idealised Wright-Fisher population that were to undergo an identical amount of genetic drift as the population of interest. Often simplistically, it is taken as the number of breeding individuals. N_e for mtDNA and the Y chromosome is expected to be one quarter of any autosome, and one third in the case of the X chromosome. This is particularly meaningful as this entails a lower diversity and that the random variation in allele frequencies by chance alone (i.e. genetic drift) is more pronounced for them. This means that alleles in a population with a small N_e are much more likely to reach fixation (frequency = 1) or become lost (frequency = 0) than for one with a larger N_e, given the amplitude of the variation in their frequency is inversely proportional to the size of the population.

To sum up, the lack of recombination and exclusive inheritance of the mtDNA through females allow to reconstruct its true phylogeny. And using the molecular clock theory (section 2.2.2), it is possible to estimate the coalescent events leading to its root.

2.1.2. The Y chromosome, looking through the male perspective

Like mtDNA, the Y chromosome is inherited in a haploid fashion but only through males. It belongs to the XY sex-determination system; the Y chromosome most commonly determining the sex in mammals via the *SRY* gene (Kashimada and Koopman 2010), initiating virilization together with others factors (Sekido and Lovell-Badge 2008). Whilst the mtDNA genome can be consider rather uncomplicated in its architecture, it is quite different for the chrY, owing to the way it came to be. The human sex chromosomes started as a pair of autosomes, where the proto-chrY acquired the *SRY* gene, followed by a multitude of events, including inversions, losses of material, fusions and duplications (Bellott *et al.* 2014; Cortez *et al.* 2014; Graves 1995) over the past 180 Myr. This tumultuous evolution resulted in a sexual dimorphism of sort, with the X chromosome and its ≈156 Mb compare to the ≈60 Mb of the chrY. This degeneration was the consequence of inversions effectively making recombination with the X chromosome impossible (Lahn and Page 1999), save for two regions, one at both telomeres of the chrY. These regions still exhibit sequence homology, allowing crossing over with the X chromosome during meiosis. The region between these two is called the male-specific region of the Y chromosome (MSY). The MSY is split into two segments, one of heterochromatin (highly condensed, transcriptionally inert) of variable length, and the other of euchromatin (extended conformation, transcriptionally active) of roughly 23 Mb. The euchromatin comprises three main classes of sequences

(Skaletsky *et al.* 2003): the X-degenerate class (XDG), the X-transposed region (XTR) and the ampliconic regions that are composed of intra-chromosomal repeats of high sequence similarity. These latter regions pose a difficult challenge when using Next Generation Sequencing (NGS), as it is virtually impossible to correctly assign the sequences and variants to their actual place. That is why NGS studies usually use a combination of less problematic regions, totalling on average around 10 Mb of sequence ((Karmin *et al.* 2015; Poznik *et al.* 2016) for instance, but see (Helgason *et al.* 2015) for an impressive 21.3 Mb).

Whilst the chrY shares the haploid inheritance and non-recombination features of the mtDNA, its very complexity made its early study in the phylogenetic context less easily tractable and bias free, as I shall briefly touch upon in section 2.3.

2.1.3. The autosomes, recombination unlocked, genealogies to 11

As we saw in the two previous sections, the mtDNA and the Y chromosome are haploid loci that do not undergo recombination. These features make reconstructing their phylogenies easier (more about it below) and also provide a gender based phylogeny. But these very features also represent their major inherent limitations. Their resolved phylogenies represent only one peculiar sex-specific realisation, or one genealogy out of many that are compatible with the demographic history for this single locus, locus furthermore smaller than the autosomes, drastically so for mtDNA. On the other hand, the autosomes can be seen more akin to a patchwork, with considerably more loci, independent or not. Combining these loci is a powerful tool for inferring the demographic history of populations as a whole, not only through the male or female prism. Although nowadays sequencing the whole genome is becoming more and more prevalent (see (Mallick *et al.* 2016; Pagani *et al.* 2016) for seminal studies in a population genetics framework, a far cry from the wealth published in the medical field), earlier efforts focused on genotyping SNPs. The International HapMap Consortium has been a leading force in providing information on variation across the human genome. The phase I of the project set to genotype at least 1 common SNP every 5 kb in 269 samples from four geographically diverse populations (Yoruban, Japanese, Han Chinese and Americans from European ancestry) (International HapMap 2005), expanding it to a total of over 3.1 million SNPs genotyped in 270 individuals two years later (International HapMap *et al.* 2007). They also genotyped an additional 7 populations, data known under HapMap 3 (International HapMap *et al.* 2010). Despite these efforts, samples of Asian origins were still underrepresented, an issue that was addressed by the HUGO Pan-Asian SNP consortium, by generating a database comprising 1,719 unrelated individuals from 71 Asian populations (Ngamphiw *et al.* 2011). However powerful these datasets may be, they still report only partial infor-

mation about the totality of the genome, potentially leaving room for a bias. One way to free oneself from this downside would be to sequence the entire genome. That is precisely what the 1000 Genomes Project (1KGP), launched in January 2008, aimed to do by sequencing at least a thousand individuals from various ethnic groups. The results of the pilot phase were published two years after (Genomes Project *et al.* 2010), the whole 1,092 genomes in 2012 (Genomes Project *et al.* 2012) and the completion of the project in 2015 (Genomes Project *et al.* 2015), reporting 2,504 individuals from 26 population, including an integrated map of structural variation (Sudmant *et al.* 2015).

Using genome-wide approaches or even whole genomes allowed to study the structure and apportionment of many loci, revealing more complex population structures ((Behar *et al.* 2010; Jakobsson *et al.* 2008; Li *et al.* 2008) amongst early studies for genome-wide and the aforementioned (Mallick et al. 2016; Pagani et al. 2016) for whole genome) than attainable by simply the study of mtDNA and/or the Y chromosome. A few relevant examples of their conclusions, in agreement or not with that of haploid genomes, shall be given in some of the following sections.

2.2. Bits of practical tools

Having treated succinctly the different parts of the human genome and some of their respective advantages and drawbacks, it is time to turn my attention to how to use and manage the information embedded in them. How do we formulate the hypothesis(-se) we are interested in testing? Indeed, the manner in which we address the present-day diversity constrains them. One manner is descriptive, comparing the distribution of diversity amongst species or populations within a species; the other is inferential, resorting to the use of explicit or implicit models of the evolutionary processes having given rise to the diversity observed. Although they form a different conceptual approach, one can easily lead the other and they are often mingled.

I shall briefly describe some of the methods that are particularly relevant to the better comprehension of the References constituting this thesis, along few others, with a clear emphasis on the mtDNA perspective. As such, several methods that are prominent in the whole-genome era, as the *F*-statistics for instance (reviewed in (Peter 2016), are not discussed here, and thus the non-exhaustive character of the next sections has to be kept in mind.

2.2.1. “S’il vous plaît… dessine-moi un arbre !” “Hein !” (adapted from Saint-Exupéry)

As the saying goes, a picture is worth a thousand words. Truly, staring at a matrix of distances or at a table of 0s and 1s might not to be the most evocative nor an endeavour readily explicit, save for some mysterious cases of illustrative

synaesthesia. Hence, representing some particular molecular results under the form of phylogenetic trees can be an attractive alternative.

2.2.1.1. Concise manual for the budding phylogenetic arborist

Phylogenetic trees represent the relationships between the different taxa, usually located at the tips of the tree. Taxa are also known in a more formal manner as Operational Taxonomic Units (OTUs). There are two types of trees: they can be either rooted or unrooted. One feature of phylogenetic trees that has a particularly important influence on how they are inferred is how the number of possible trees evolves according to the number of taxa. It is simply staggering; for t taxa there are (2t-5)!! unrooted trees and (2t-3)!! rooted ones. In less abstract form, with 10 taxa for instance, there are 2,027,025 unrooted trees compared to 34,459,425 rooted trees. Penny et al. (Penny, Hendy and Holland 2007) give the example of how long it would take to calculate all the trees for 20 taxa, assuming that we can calculate one million trees per second and that each calculation step would be equal regardless of the algorithm. There are $\approx 2.2 \times 10^{20}$ trees, $\approx 3.16 \times 10^{7}$ seconds per year, so we can calculate $\approx 3 \times 10^{13}$ trees per year, meaning it would take ≈ 7 million years to calculate all the trees. It is therefore evident that not all possible trees can be inferred. We shall mention in passing that some heuristics searches have been developed to circumvent that problem, the limited (local) or greedy searches and the hill-climbing and related ones (Penny, Hendy and Holland 2007). Another crucial property of the phylogenetic trees is that as time (or evolution for that matter) goes on, the branches keep splitting, they never merge together or even simply disappear. This clearly is in contradiction with various biological processes taking place (recombination, migration, parallel mutation, gene transfer, extinction…). Networks are methods that readily incorporate in their framework ways to accommodate some of these in their structure, via reticulation (see below).

Regardless of their type, data used to draw trees/networks fall into two categories: distances and characters. Stating the obvious, in the former category, genetic distances have to be calculated first, resulting in a distance matrix. Regarding the character approach, these characters are discrete units of evolution, be it SNP, copy-number variation (CNV) like in the case of microsatellite repeats or simply phenotypic traits of interest.

Dichotomy anew, there exist also two main classes of phylogenetic tree construction methods. The first class is the clustering methods that, aptly named as they are, cluster taxa together in a hierarchical way. The Unweighted Pair-Group Method with Arithmetic Mean (UPGMA) is one of most well-known clustering methods, along with the Neighbor-Joining (NJ) one. The second contains the searching methods, which in theory explore the whole range of possible trees, such as the Maximum Parsimony (MP) or Maximum Likelihood (ML). However, as aforementioned, it quickly becomes computationally impossible to explore all the trees given the extremely large tree space. Therefore,

they explore representative subsets of the tree space, finding the best tree for each of these subsets and finally comparing the different trees, as alluded above.

It is essential to state clearly the specifics of the method applied: what is the optimality criterion, that is a measure by which to evaluate how a given tree "fits" the data; what is the search strategy for finding the optimal tree(s) and what are the assumptions regarding the mechanisms of evolution underlying the data (Penny, Hendy and Holland 2007). Besides these essential combinations, each method should have five desirable properties: efficient, consistent, powerful, robust and falsifiable (Penny, Hendy and Steel 1992). The efficient property refers to how quickly the tree is constructed. The consistent character of a method pertains to its capacity to recover the same tree if more data are added. A method will also be inconsistent if the assumptions on which it is based (i.e. the mechanisms of evolution) do not apply to the data. A method will be deemed powerful if it can get to the correct tree with scarce data (number of taxa or sequence length). Robustness could be seen as an extension of the previous property. A method would be robust if it is impervious to minimal (or more consequent) deviations from its assumptions, that is not becoming inconsistent. Last, but not least, the epitome of Popperian thought, it is possible to test the validity of the assumptions under which the method is based, namely that "the data, in principle, must be able to falsify the model" (Penny, Hendy and Holland 2007). Unsurprisingly, not a single method does possess all these properties. Some methods are inherently fast at the expense of other properties, and it is obvious that navigating through large tree spaces is taxing timewise; these subtleties shall be kept in mind when choosing a method.

Another subtlety, of semantic nature this time, regards parsimony. As explained in (Penny, Hendy and Holland 2007), there are two main meanings implied by parsimony. The first is the Principle of Parsimony, exemplified by Occam's Razor, which stipulates that the simplest hypothesis amongst competing ones explaining the data should be favoured. The second is that one should minimise the amount of mutations (or steps) when building a tree. These meanings need not be mutually exclusive but can be so at times, i.e. the least number of mutational steps may be the simplest of explanation but not the correct one.

Whilst MP looks at minimizing the number of evolutionary changes, ML takes the trees and calculates how likely they are to give the data. Hence, under a specific evolutionary model, the tree with the maximum likelihood of producing the data will be the best tree. There is also an approach that changes the way to look at the terms of the probabilities in the form of the Bayesian methods. They determine the probability of the tree given the data, making use of Markov Chain Monte-Carlo (MCMC) simulations to calculate the parameters. MCMC are algorithms used for sampling from a probability distribution, in this case from the posterior distribution. In a more befitting form, with H referring to hypothesis, O to outcome and P to probability, the ML methods calculate P(O|H) and the Bayesian ones P(H|O).

However, as indicated at the beginning of this section, the tree(s) resulting for these methods are acyclic. When dealing with mtDNA HVS-I sequences for instance, parallel mutations are something very real, and have to be taken into account. Phylogenetic networks readily address these kinds of issues. There are once again various methods available (see (Huson, Rupp and Scornavacca 2011) for more details), but we will focus briefly on the ones more germane to mtDNA studies. First are median networks, which contain all most parsimonious trees (Bandelt *et al.* 1995). Yet, with the increase in samples size, so does the dimensions of the cubes of reticulations. The authors proposed a procedure to reduce this by using the frequencies of the mitochondrial haplotypes combined with a compatibility argument, thereby reducing the complexity. The other method, median joining (Bandelt, Forster and Röhl 1999), draws on the previous one whilst solving some previous issues like accommodating multistate characters and speed of calculation. Both methods have been widely applied for reconstructing mtDNA phylogenies.

As a non-sequitur of sort, the influence of the reference sequence should be mentioned. With character state methods, its choice can have an influence on the actual evolutionary state of the character. As touched upon when introducing the mtDNA sequence, the rCRS coalesces to the European hg H2a2a1 (van Oven and Kayser 2009), that is phylogenetically not the ancestral sequence to all mtDNAs, hence when scoring mutations along the branches, not all of them truly reflected an ancestral to derived "transition". Also, it *de facto* misplaced the true root of the mtDNA phylogenetic tree. This issue of a true ancestral reference transcends the mitochondrial centric field of population genetics. The chrY reference sequence (Skaletsky et al. 2003) is from one individual, but with a 49,XYYY karyotype (Foote *et al.* 1992; Tilford *et al.* 2001), and it belongs mostly to hg R, with portions belonging to hg J. In the case of the human reference, the GRCh37 is a mix of 13 different individuals (Kidd *et al.* 2010).

Regardless of the reference, once trees or networks are obtained, there is one way to add an extra layer of information, that is the geographical origin of the taxa.

2.2.1.2. Phylogeny + geography = phylogeography

A wonderful example of a portmanteau, the term phylogeography was coined in 1987 by Avise et al. (Avise *et al.* 1987) and is the field of study "concerned with the principles and processes governing the geographic distributions of genealogical lineages […]" (Avise 2000). Most often, the analyses rely on the visual inspection of frequency maps. There have been attempts to approach geographic patterns by methods with a quantitative focus, such as the nested cladistic analysis of Templeton (Templeton 1998) or the founder analysis of Richards et al. (Richards *et al.* 2000), of which an earlier incarnation (Richards *et al.* 1996) drew a slew of acrimonious comments (see Box 12.3 p. 382 in (Jobling *et al.* 2014a) for an expanded view). A critical assessment of the value

of phylogeographic methods was also published by Rasmus Nielsen and Mark Beaumont (Nielsen and Beaumont 2009). Their central issue is that most phylogeographic studies assume that it is possible to interpret branches of phylogenetic trees as an actual evidence that some specific historical demographic events in a geographical context have occurred, that is that "that ancestral history can be directly deduced from estimated gene trees". They contrast it with the theoretical population genetic approach by which "gene trees are random outcomes of stochastic population level processes". They caution about using the former reasoning for several reasons. One is the crucial influence sampling has on the resulting gene tree; depending on the samples, the resulting trees could be quite different. A corollary to this fact is that the structure of the tree is influenced by which individuals left descendant(s), or not, in the next generation. And in that regard, it can be misleading to interpret this specific structure as more due to a particular ancestral demographic event than the random process by which offspring compose the next generation. They further stressed that the distribution of gene trees in populations that have otherwise experienced the same demographic history (identical N_e, same geographical distributions) could be extremely different. This is compounded by the fact that multiple demographic models may fit a gene tree equally well. They however note that if one were to assume that "population migration events are always associated with population bottlenecks", thereby increasing the possibility of the monophyletic character of the resulting population, and in the absence of no subsequent gene flow, then it might be possible to assume that gene trees reflect directly population history.

It remains that, when using a fully resolved and unambiguous phylogeny of mitogenomes (full sequences), there are evident and varied geographical patterns emerging according to the level of phylogenetic resolution (see Figure 2 for an attempt of ostensive definition). It is common to broadly divide the mtDNA phylogeny into its main macrohaplogroups L, M and N, and their respective frequency map might explain why, especially in the case of macro-Hg L and M. L lineages are mostly found in Sub-Saharan Africa, where they frequently encompass the entire mtDNA genetic pool of the populations inhabiting there. Macrohaplogroup M, despite having a wider transitional region, is for his part very frequent in South and East Asia. In actuality, if one combines frequencies and phylogeny (and its by-product diversity), one can tentatively posit that such and such hgs or sub-clades originated from a very localised (or not) region. It is easier to do so if the hg or sub-clade in case has a fairly restricted dispersal area and is nested within another one with itself a clear geographical spread.

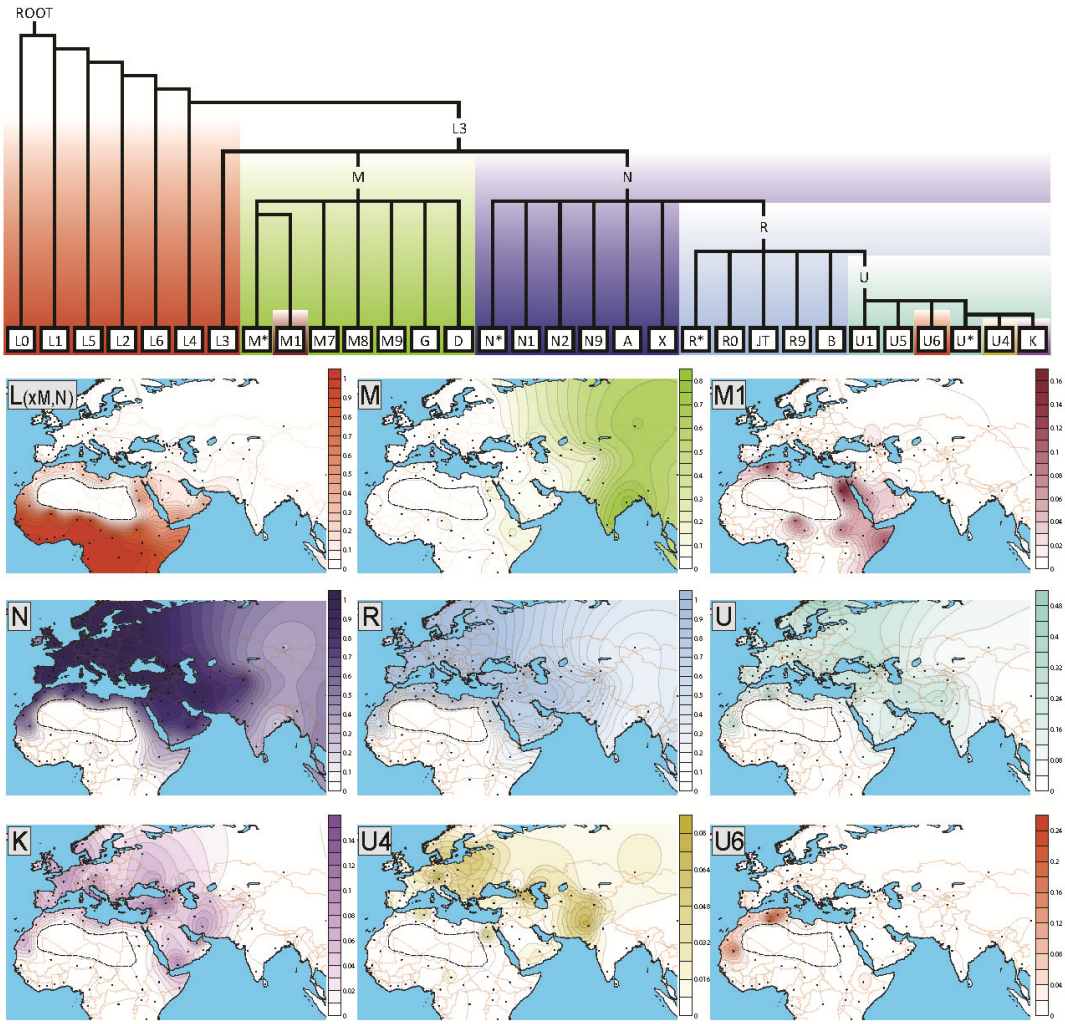

Figure 2. Visual explanation of phylogeography. The upper part represents the phylogenetic tree of mtDNA as currently resolved (inspired and modified from phylotree.org (van Oven and Kayser 2009)). The boxes behind the branches follow the same colour codes as the frequency maps on the lower part. Only the frequency bars for the maps of macrohaplogroups N, R and all the L lineages [denoted L(xM,N)] go from 0 to 1 (that is 0 to 100%), the others are individual to each macrohaplogroup/Hg. The maps are inspired by figure S1 from (Olivieri *et al.* 2006) and based on a revised and expanded data set drawn from Table S1 of the same publication. The dash line is the contour used to represent the break "line" of the Sahara.

The phylogeographic approach is frequently used despite the caveats, limitations and other considerations mentioned above. We shall indirectly address this in section 2.3.

2.2.2. Molecular clock

Linus Pauling, besides being awarded twice a Nobel prize (in 1954 in chemistry for his research on the nature of the chemical bond and then in 1962 for Peace), was a chemist and biochemist (biochemist ⊆ chemist). He discovered, together with Émile Zuckerkandl, the first evidence of what would give rise to the molecular clock theory by studying the protein sequences of haemoglobin and cytochrome c in different species (Zuckerkandl and Pauling 1962). The idea of a protein molecular clock was formalised by Margoliash (Margoliash 1963) and the actual term of "Molecular evolutionary clock" introduced in a second publication by Zuckerkandl and Pauling in 1965 (Zuckerkandl and Pauling 1965). The ticking of the clock is actually the accumulation of mutations, one of the main evolutionary force, and the idea that a clock may be ticking at similar speed in different and distant organisms (that is these diverse lineages accumulate mutations at a similar rate) did not sit too well with many other scientists (see (Kumar 2005) and references therein for instance). Yet, it provided a fertile ground for further research and ideas, one of which turned out to be rather pivotal, when Motoo Kimura developed the neutral theory of molecular evolution (Kimura 1968). Together, these ideas constituted a new and powerful tool in biology. Although, as shown later by various studies ((Kumar 2005) and (Bromham and Penny 2003) for example provide more details on that), the clock is not to be taken as ticking regularly (far from it, it is rather stochastic, with mutation rates varying), it nonetheless proves to be an invaluable method for testing a large and varied horizon of hypotheses.

Indeed, being able to situate diversity and events into a temporal frame adds an essential extra dimension. For calculating the human mtDNA mutation rate, two approaches have been used. The first one is based on pedigree, using parents/offspring pairs and counting the number of new mutations that emerge between generations (Heyer *et al.* 2001; Howell *et al.* 2003; Santos *et al.* 2008; Sigurgardottir *et al.* 2000). The second one is based on phylogeny, using an external calibration point, often coming from archaeological records, such the divergence time between chimpanzee-human or a well characterised human first settlement. An example of the latter is what led to the rate that was most frequently used for HVS-I. Forster and colleagues (Forster *et al.* 1996), reappraising Native Americans mtDNA origins, derived their rate of one transition in every 20,180 years using a network-based method under the assumption that the expansion of haplogroup A2 occurred 11.3 KYA. But the HVS-I can be problematic, with parallel mutations (Tamura and Nei 1993), and varying substitution rates. Ingman et al. (Ingman *et al.* 2000) introduced a new rate that made use of the coding region mutations using 53 full sequences, and a human-chimpanzee split of 5 Myr. Mishmar et al. (Mishmar *et al.* 2003) on the other hand used a human-chimpanzee of 6.5 Myr and 104 full sequences to derive their rate for the coding region. Unfortunately, the influence of selection was not properly acknowledged in its calculation. As a matter of fact, even in their study it was shown that non-synonymous mutations were not occurring evenly

on the mtDNA phylogenetic tree. This class of mutation are more prevalent in the terminal branches (Elson, Turnbull and Howell 2004; Moilanen and Majamaa 2003; Ruiz-Pesini *et al.* 2004) and amongst the younger clades (Kivisild *et al.* 2006). To remedy that, a rate taking into account only synonymous mutations was introduced by Kivisild and colleagues (Kivisild et al. 2006). This rate was further refined in Loogväli et al (Loogvali *et al.* 2009). Addressing similar concerns, as well as considering the time-dependency of the mutation rate ((Endicott *et al.* 2009; Loogvali et al. 2009) and references therein), a group extended their approach to the entire mtDNA molecule (Soares et al. 2009). This issue of time-dependency was also tackled in the study of Henn et al. (Henn *et al.* 2009). Endicott and Ho (Endicott and Ho 2008) explored the effects on the mutation rate of using different partitions of the mDNA, as well as these of using multiple internal calibrations versus the usual human-chimpanzee split. There have also been studies leveraging ancient mtDNA from securely dated samples as calibration points, either using mtDNA from diverse hgs (Fu *et al.* 2013) or specifically Neolithic samples belonging to hg H from Central Europe (Brotherton *et al.* 2013).

It is important to mention one potential crucial caveat regarding the use of the rho summary statistic (ρ) introduced in (Forster et al. 1996), which is frequently used for converting most of these rates into actual dates. This scalar is not very robust to cases of deviation from idealised assumptions, such a constant population, or can be biased in the presence of natural selection and other rates variation (Cox 2008). Whilst its use remains nonetheless the standard approach in many mtDNA studies, this potential lack of robustness should be kept in mind.

Controversies surrounding the mutation rate of mtDNA also exist for the chrY. Classically, when researchers were limited to typing only a handful of SNPs, they used a set of Short Tandem Repeats (STR) to calculate the coalescent estimates, but then they had to choose between using the "genealogical" rate measured in families/pedigrees (Heyer *et al.* 1997) or the evolutionary one (Zhivotovsky *et al.* 2004), the latter being threefold slower. With chrY sequencing data, these two types of rate still exist, but they are much less further apart than for STRs, with the proposed use of a consensus envelope rate ((Balanovsky 2017) and reference therein).

2.2.3. Coalescent theory and Bayesian Skyline Plot

The coalescent theory is a mathematical theory, based on several stochastic models, that was introduced in the early 1980s. It is mainly attributed to Kingman (Kingman 1982a, Kingman, 1982 #2320; Kingman 1982b), who recounts its inception in (Kingman 2000), but it was also independently discovered by Hudson (Hudson 1983) and Tajima (Tajima 1983). One of the models on which it often relies is the Wright-Fisher model of neutral evolution. Further, it stems from the realization that it is generally easier to reconstruct the gene genealogy

of a population backward in time rather than forward in time, as forward simulation would require to know all the sequences present in the current population (Jobling *et al.* 2014b; Nordborg 2001). The coalescent approach is based on two fundamentals insights. The first is that, if we assume variants to be selectively neutral, we can separate the genealogical process from the mutational one. We can simply reconstruct the gene genealogy and later on drop the mutations along the branches according to the mutation process and their probability to occur between parents and offspring. The second insight regards the modelling backward in time of the gene genealogy itself. It can be done so for only a group of individuals, not for the population as a whole (see Figure 3).

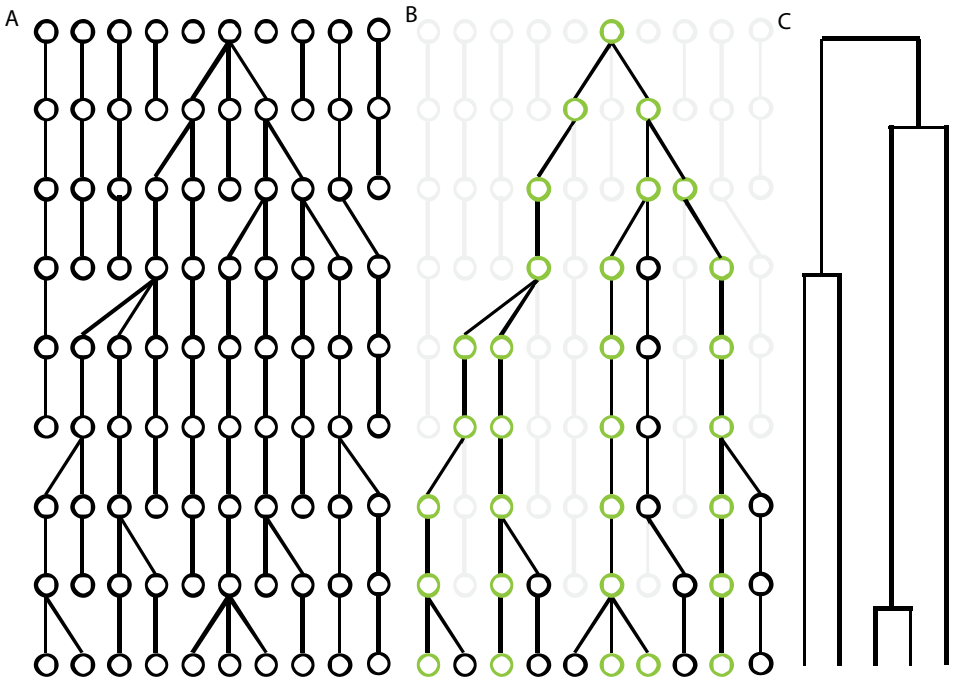

Figure 3. Coalescent approach. (A) Schematic representation of the genealogy of a population of constant size 10 through 9 generations. (B) The actual genealogy of the extant lineages. Even sampling only some samples from the whole population (green circles) gives access to its TMRCA. (C) The coalescent tree of the lineages sampled. (Figure freely based on Figure 6.12 from (Jobling et al. 2014b))

As we move backward in time, some lineages will share a same ancestor, i.e. they coalesce. The coalescent approach allows the description of the influence population genetic processes have on the shape of the genealogy of sampled gene sequences (Drummond *et al.* 2005). Several methods based on it have been developed to estimate different population genetics parameters under

various biological scenarios. One interesting estimate is N_e and its change through time. N_e can be measured as the rate of coalescence. The genealogy of a population of constant size and those of populations under expansion or contraction recently or since more ancient times are different. These differences are reflected in the distribution of the number of coalescent events through time. Analysing the variation of the rate of coalescence through time gives access to N_e and there are several methods developed to do so.

The Bayesian Skyline Plot (BSP) approach is one of them and reconstructs past population dynamics using the framework of the coalescent theory (Drummond et al. 2005); the method is implemented in the software BEAST (Drummond and Rambaut 2007; Drummond *et al.* 2012). It draws on the variable population size coalescent model and is a refinement of the skyline plot introduced earlier (Pybus and Rambaut 2002). The skyline plot model was introduced to address the issue of the choice of the correct "demographic model". A wrong choice could lead to biased and fallacious estimates. Using sequence data and a MCMC sampling procedure, the BSP gives the posterior distribution of effective population size through time. This addresses another issue which still affected the general skyline plot, which was taking into account the errors pertaining to reconstructing the phylogeny, as the estimates in the skyline plots methods were inferred on an estimated genealogy, not from the sampled gene sequences. The BSP not only incorporate these potential errors, it also includes the uncertainties inherent to the stochastic nature of the coalescent. Yet, a more fundamental assumption of the method, and indeed of the coalescent theory itself, might be often violated. As an extension of the Wright-Fisher model, a panmictic population is assumed, and applying the BSP to structured populations can lead to spurious results (Heller, Chikhi and Siegismund 2013), albeit in some specific cases.

2.2.4. Analyses for revealing demographic history from autosomal data

The first few methods are not strictly applied to autosomal data only. Many of them were indeed developed well before the advent of autosomal data sets, and are staples of many classical human population genetics studies. A crucial point to bear in mind is that methods rely on assumptions about the biological processes that begot the data, and as such, if the actual processes underlying them depart from these assumptions, their interpretation is bound to be biased or could lead to erroneous inferences.

2.2.4.1. Nei's standard genetic distance *D* and F_{st}

Two frequently used methods of calculating genetic distances are Nei's standard genetic distance *D* and F_{st}. Nei's standard genetic distance *D* associates the probability of drawing two identical alleles from two different populations with the probability of drawing identical alleles from the same populations. Its values range from zero to infinity. F_{st}, a member of the fixation indices family, measures population differentiation caused by genetic structure by comparing the genetic diversity found within two subpopulations to the genetic diversity of the total population. F_{st} values vary from 0 to 1; 0 being no genetic differentiation between the subpopulations and tending towards 1 when they are highly differentiated. The values are often presented in the form of a heat map, where it is visually easier to distinguish the substructure(s) revealing the potential subpopulations.

2.2.4.2. Principal Component Analysis, one of the go-to analyses

The Principal Component Analysis (PCA) is a statistical procedure that belongs to the multivariate type of analysis. It extracts principal components, also known as eigenvectors, one after another in such a way that they contain the largest possible uncorrelated variance present or remaining. As such, the first PCs are usually the ones plotted and shown, as they contain the highest individual variance. PCA is nowadays usually performed using the covariance matrix of the genome-wide genotyping data, but it has also been applied to allele frequency data, such as mtDNA haplogroups frequencies. Although the method aims at extracting the maximum variance for each PC, in the case of genome-wide data, the overall variance explained by the first two or three PCs is frequently quite low. That can be intuitively understood by the total amount of "data point" included in the dataset itself, with, for instance, upwards of 195,000 SNPs for each sample in (Novembre *et al.* 2008). Also, if the loci used are highly variable between the populations, the variance explained will likely be higher. It is important when applying PCA to genome-wide data to take into account the potential severe bias from the effect of linkage disequilibrium (LD). If SNPs in LD are not removed prior to the analysis, some of the PC may actually rather reveal long LD blocks than the actual sub-structure of interest; but that this can also be taken advantage of (see the use of co-ancestry matrix with fineSTRUCRURE in section 2.2.4.3). Sample sizes can have an impact as well; if one or more populations in the dataset are overly represented, it is their overall variance that will likely drive the loading of the eigenvectors, not that of the whole set (McVean 2009). To note also, whilst visually attractive, with possible clusters of discrete populations apparent, interspersed with admixed ones, PCA plots are not formally indicative of populations' movements. Indeed, PCs have been used to generate synthetic maps, one most prominent example found in Cavalli-Sforza et al. (Cavalli-Sforza, Menozzi and Piazza 1994) with

the map of the first PC in Europe which has been used to support the hypothesis of the Neolithic demic diffusion. It has been shown ((Francois *et al.* 2010; Novembre and Stephens 2008)) that such gradients may arise inherently to the method.

A closely related method, Multi-Dimensional Scaling (MDS), is also used sometimes, but it is conceptually different in that it aims at conserving the distances present in the dataset in the lower dimensions onto which they are represented.

2.2.4.3. Stacked charts, creative palettes, the realm of cluster analyses

Cluster analysis underlies most popular admixture algorithms, grouping objects, or individuals in this case, into separate clusters. Amongst the algorithms available, the first described was STRUCTURE developed by Pritchard et al. (Pritchard, Stephens and Donnelly 2000), using a Bayesian framework, with later refined programs FRAPPE and ADMIXTURE introduced by Tang et al. (Tang *et al.* 2005) and Alexander et al. (Alexander, Novembre and Lange 2009) respectively. They expended the capacities of STRUCTURE, by notably using maximum likelihood approaches instead of MCMC. They are dealing with "global ancestry estimation", that is they estimate what is the ancestry proportion from each contributing population, but this proportion is taken as an average across the whole genome. The modelling assumptions are that the populations are in Hardy-Weinberg equilibrium and that there is complete linkage equilibrium between the loci within the population. As such, the programme will attempt to define groupings into population sample which maximize Hardy-Weinberg equilibrium and minimizes LD. Based on their genetic similarity to the predefined *K* different clusters, the individuals will be assigned varying proportions of them. An individual can show a mosaic of the different *K*-s or sometimes only one of them. The number of *K*-s is predefined by the researcher, so one issue is which model (what number of *K-s*) best describe the genetic structure in the sample. One could use prior knowledge (or an educated guesstimate) regarding the putative number of postulated "ancestral" populations that gave rise to the extant populations of interest or simply run the algorithm for a range of *K* clusters and calculate the fit of the different *K-s* to the real data. Different approaches have been used to calculate the fit (Alexander, Novembre and Lange 2009; Evanno, Regnaut and Goudet 2005; Pritchard, Stephens and Donnelly 2000), although a rigorous estimation of *K* represents a statistical challenge (Alexander, Novembre and Lange 2009; Falush 2016). Of course, whether a biologically significant number of clusters does exist needs to be kept in mind. One of its corollary is that interpreting the plots in terms of *K* true ancestral populations can likely become an over-interpretation. Another cautionary note in the interpretation of the ubiquitous bar plots is that the models on which they are based may "missed" more complex demographic

scenarios. For instance, Falush et al (Falush 2016) discuss three qualitatively different scenarios that give virtually identical ADMIXTURE plots when focusing on 4 populations out of 12, with K=11. In the first case, ADMIXTURE correctly reconstructs the recent admixture of population P2 as being a mix of the other sampled populations P1, P3 and P4. But in the second case, it did not correctly reconstruct the scenario where population P2 was a mix of population P1 and one unsampled one (termed "ghost"), this ghost population being closely related to population P3. In the last scenario, ADMIXTURE fails to properly infer a very recent bottleneck, where population P1 is a sister population of P2, inferring a component unique to P1. That is something that can frequently happen in case of a population that has experienced strong genetic drift. It will quickly "receive" its own component, which in turn will be "sprinkled" to many other individuals. Beside these potential pitfalls, there is as well a more global issue about these STRUCTURE-like programs. They do not take any advantage of the position of the SNPs they are analysing. Indeed, one of the firsts steps in the pipeline is to remove SNPs in strong linkage disequilibrium; not doing so seriously impact the results.

Exploiting the patterns of haplotype similarity, that is using SNPs that are in strong linkage disequilibrium allows reaching unprecedented finer levels of population structure (Lawson *et al.* 2012). Roughly, each individual in the data set has its chromosomes reconstructed by copying chunks from the other individuals. This process is referred to as "chromosome painting", and has been implemented in CHROMOPAINTER (Lawson et al. 2012). The results can be summarised in a co-ancestry matrix that encapsulates practically all the information used by PCA and STRUCTURE-like programs which treat markers independently. Conversely, when the LD information is taken into account, fine scale population structure can be distinguished by running the program fineSTRUCTURE on the co-ancestry matrix calculated by CHROMOPAINTER. Another powerful extension to the latter was developed by Hellenthal et al (Hellenthal *et al.* 2014) with GLOBETROTTER, a program able to identify, date and describe potential admixture events that could have occurred within the last ~ 4.5 KY.

2.3. Why, how why?

After all these prolegomena, one would be well placed to enquire: "What is this all about?". Indeed, stringing together facts after facts, methods after methods, theories after theories does not necessarily render any clearer what is this all about. It all pertains to population genetics studies, where we are studying, as its name implies, populations, and in our specific case, human populations. And even more restrictively here through primarily mtDNA and its reconstructed phylogeny. But it is actually inscribed within the more global framework of evolution, where technical advances together with new ideas and theories allowed (and continue to do so) refining our understanding of how it acts and has shaped the biological world as we observe it. Ideas and theories raise

questions, and as alluded in the introduction, and here and there above, we seek to answer some of these questions by studying mtDNA.

As an example of the usefulness of mtDNA and its phylogeny, before later on expanding our scope, we may recount how it came to shed some light on a central question that was hotly debated: where did *Homo sapiens* originate? Paleoanthropologists and anthropologists were arguing over two main competing hypotheses: the multiregional hypothesis (Wolpoff, Wu and Thorne 1984) and the Out of Africa one (or "Noah's ark" at its inception (Howells 1976)) (Stringer and Andrews 1988); OoA for short. The former proposed that the human species arose around two million years ago and that the transition from *H. erectus* and Neanderthals to *H. sapiens* took place in different places. In the latter one however, this transition took place in Africa and it is these humans that then went on to colonise the rest of world, replacing the existing hominins there. Enters genetics.

Whilst mtDNA sequence had already been resolved (section 2.1.1), sequencing in the early 80's was not as easy as it is nowadays, and the standard way of studying mtDNA diversity was by doing Restriction Fragment Length Polymorphism (RFLP) analysis. Some of the early studies include the one from Wesley M. Brown in 1980 (Brown 1980), which already extolled the possibilities offered by such method, stating "[…] this method of analysis makes possible the investigation of many questions concerning human population genetics, evolution, and recent history". Interestingly, one of its conclusion was that *Homo sapiens* could have experienced a serious bottleneck event "as recently as 180,000 years ago". It also introduced the concept of morphs, that is the patterns obtained after digestion by endonuclease enzymes, an ancestor of kind of the later haplotypes routinely analysed now. This study was quickly followed by that of Denaro et al. (Denaro *et al.* 1981) using 235 individuals, classified as Caucasians, Orientals and Africans. From the 6 morphs revealed and their respective frequency, they suggested that Asia played a central role in the radiations of human ethnic groups. But probably the study that had the biggest impact of the multiregional versus OoA debate is that of Rebecca L. Cann and colleagues published on 1 January 1987 in Nature (Cann, Stoneking and Wilson 1987), which led to the famous front cover of Newsweek of 11 January 1988, depicting an Adam and Eve of African appearance in the Garden of Eden. Studying 147 individuals, they reconstructed a phylogenetic mtDNA tree using the parsimony principles. The tree shown two major branches, the most divergent one (hence most ancient) comprising only African samples. The conclusion drawn from the tree structure was that modern human populations had an African origin.

Although relying on RFLP only, the basic structure they found, whereby the deepest and oldest branches are restricted to Africa, has stood the test of time (see (Behar *et al.* 2008)). Throughout the following years, the phylogenetic tree of mtDNA has been improved. After being "restricted" to RFLP, researchers started to use a combination of RFLP and sequencing HVS-I ((Torroni *et al.* 1993a; Torroni *et al.* 1993b) among the early ones), to finally simply

sequencing the whole mtDNA (Ingman and Gyllensten 2001). But this continuous improvement was often not the sole aim of the studies, if at all. It came together with addressing wider questions, such as, amongst others, the peopling of Americas and the relation of Native Americans with aboriginal Siberians in the two Torroni *et al.* references just cited. The power of mtDNA and its phylogeny has been, and still is, applied in different ways. It can be used to describe in a general way a population or populations of interest and by placing the resulting gene pool in a broader context, we can answer whether or not some potential migrations occurred, via the sharing of identical haplotypes for instance. In that particular case, using full sequences leaves less room for criticism than more intermediate level of phylogenetic resolution. It is also possible to focus only on some specific hgs, and, through molecular date estimates, to connect them to major climatic events like in the case of the Last Glacial Maximum (Torroni *et al.* 2001). These are simply but a few illustrative examples, and they are equally valid regarding the chrY, revealing then the genetic legacy of the male lineages. And by combining both, it is possible to see if there exists some discordance between them. An elegant example of this possibility is the study by Karmin *et al.* (Karmin et al. 2015) that shown a recent and strong bottleneck of the chrY lineages, which is not seen in mtDNA. One can also reveal the different influences leading to the contemporary make-up of the male and female gene pools, as in (Chaubey *et al.* 2017; Silva *et al.* 2017) for recent examples focusing on South Asia.

It is important to mention that whilst one of the tenets of phylogeography is that we can decipher ancient movements by studying extant populations (refer to section 2.2.1.1 for a critical view about this claim), there exists also a more straightforward approach, namely to directly study ancient DNA. In that regard, mtDNA is particularly suitable given the number of mtDNA copies present in a cell and was therefore the target of choice in the infancy of the field of what can be called archaeogenetics (although the number of samples to draw conclusion on were low in the beginning). To cite just a few major discoveries from a mtDNA perspective, we may start with the studies by Krings and colleagues. They sequenced the entire HVR-I of the Neanderthal-type specimen (Krings *et al.* 1997), followed by the HVR-II few years after (Krings *et al.* 1999). They showed in the first study that the Neanderthal sequence fell outside the variation of modern humans. And as the estimated common ancestor of Neanderthal and modern human mtDNA was four times that of the common ancestor of human mtDNAs, they suggested that "Neandertals went extinct without contributing mtDNA to modern humans", a suggestion confirmed by the results of the second study (but more about Neanderthal and its legacy in modern humans below). As a second example, recall the study of Richards et al. (Richards et al. 2000) cited in section 2.2.1.1., which reached the conclusions that the majority of the mtDNA lineages found in today's European populations entered Europe via several waves during the Upper Palaeolithic and that less than one-quarter can be related to a Neolithic influx. They were to some extent confirmed by the studies of Haak and colleagues (Haak *et al.* 2005; Haak *et al.* 2010) who studied

samples associated with the earliest farming culture in Europe, the Linear Pottery Culture or LBK. They found that many of the early Neolithic farmers had mtDNA lineages that are related to modern-day Near-East and Anatolia, but that most of these lineages are now almost absent from modern European, thus favouring a Palaeolithic ancestry for them. Yet, this is where we reach a decisive turnaround, one in which NGS has played a pivotal role. By its very principle of amplifying small fragments of DNA, that technology proved to be particularly suitable for studying ancient DNA, and not simply mtDNA, but the actual entire genome; quality of the preservation allowing. To say that it led to a slight change in how we view our past history would be a prime example of understatement.

I shall not give a comprehensive description (see (Skoglund and Mathieson 2018) to this end), but highlight only a few that are a relevant segue to the ones above. In that regard, if the sequencing the first one million base pairs Neanderthal DNA (Green *et al.* 2006) was a great achievement, the draft sequence based on three individuals (Green *et al.* 2010) had a profound implication on the OoA. Whilst it was believed that modern humans left Africa and went on to colonise the rest of World, strictly replacing the archaic hominin, some of the analyses' results showed that Neanderthals shared more genetic variants with present-day humans living in Eurasia than with those in sub-Saharan Africa. This suggests that there actually has been gene flow from Neanderthals into what was the ancestors of non-Africans. And Neanderthals were not alone. After sequencing the DNA extracted from a phalanx bone found in the Denisova cave of South Siberia, the resulting archaic hominin sequence revealed that more archaic introgression took place, this time mostly found amongst present-day Melanesians (Reich *et al.* 2010). This lead Svante Pääbo to speak of "leaky replacement" as quoted in an article from Ann Gibbons in Science (Gibbons 2011). The resulting picture of the origin of modern humans has been radically transformed by these studies, and the models of the dispersal of modern humans are being further refined with the continuous sequencing of ancient individuals. Such one regards Europe and Western Eurasia. To put it ever so briefly, the present-day Europeans are not simply direct descendants of Palaeolithic hunter-gathers sprinkled with some Neolithic farmers, nor the result of a massive Neolithic wave of advance. They are the results of at least three admixture events (Lazaridis *et al.* 2014), and the finer grain picture is temporally and geographically more intricate (see (Skoglund and Mathieson 2018) and references therein).

In summary, it is possible to probe human past and recent demographic events through the study mtDNA and various *ad hoc* methods developed in a rigorous theoretical framework. It has proved to be a powerful tool that provided oftentimes robust evidence, such as in the case of the OoA hypothesis. But as is equally evident from above, it has some inherent limitations, and it is important to not forget about them when reading about results inferred from it.

3. AIMS OF THE STUDY

As described in the previous chapter, mtDNA allows to answer questions about population dynamics and history. REFs I, II and IV of this thesis use the extant genetic pool to do so whilst REF III focusses on two specific hgs.

Their respective aims are as follows:

- REF I
  - To describe the Ethiopian and Yemeni mtDNA gene pool using high-resolution phylogenetic analysis
  - To explore potential events of gene flow across the Red and Arabian Seas

- REF II
  - To characterise French mtDNA gene pool and put it in the European context
  - To explore the possible peculiarities of two more restricted part of France, Brittany and the Basque country, and to see if they may reflect some specific connections

- REF III
  - To characterise in details the phylogenies of hgs U6 and M1
  - To explore when and where their sub-clades and themselves originated to elucidate their claimed shared history
  - To put these expansion events into an archaeological and linguistic perspective

- REF IV
  - To describe the mtDNA and chrY gene pools of the Pashtuns, Tajiks, Hazaras, Uzbeks and Turkmen of the Afghan Hindu Kush and to genotype of subset of them
  - To see if the different genomes reveal a similar story
  - To ascertain whether the Hindu Kush, and more globally Central Asia, was a site from which some human migrations originated or rather through which these latter ones transited.

4. MATERIAL AND METHODS

The details regarding the origin of the samples and the methods used for their analyses are described in the respective publications forming the basis of the present dissertation.

All 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.

5. RESULTS AND DISCUSSION

In this section, I shall present separately the main results and conclusion for each of the four articles composing this thesis. After the presentation of the results, I will give a very succinct overview of selected articles published subsequently using similar type of data, if available, and which may bring a new perspective on the questions or further support the central conclusions exposed before.

5.1. Ethiopia and the continuous "Out of and into Africa" around the Bab-el-Mandeb; and beyond (Ref. I)

The then

By characterising the mtDNA pool of Ethiopians and Yemeni using a high-resolution phylogenetic approach, we sought to determine if those pools reflected the enduring connections around the Bab-el-Mandeb supported by archaeological and historical evidence, but in a more recent time frame than the OoA. We also put the observed variation into a broader geographical context, extending our comparison panel to the rest of Africa and the Near East.

On a very general level, the Ethiopian and Yemeni mtDNA compositions are broadly similar, almost half and half of hgs usually described as sub-Saharan and Western Eurasian. This is reflected on a Multidimensional scaling plot (using the F_{st} values based on hg frequencies) where they cluster together, between North West Africa/the Near East on one side and sub-Saharan Africa/ Mozambique on the other. One notable feature on the plot is that even though the Ethiopian samples were grouped according to their ethnic affiliation (reflecting the Semitic and Cushitic speaking populations), they still nonetheless clustered together. The place of the Yemeni mirrors the results of the estimated admixture proportions, with a predominant Ethiopian origin along with admixture from the Near East. However, these results do not accommodate the temporality, nor the potential more precise origins of the lineages involved. To focus solely on the last thousands of years, only lineages that were exact matches or one step-derivative were considered for a phylogeographic founder analysis.

The diversity we observed in the composition of sub-Saharan hgs was different between Ethiopia and Yemen. Concentrating on the Yemeni angle, we showed it reflected three different passages. The first was indicated by exact haplotype matches with south-eastern Africa samples, likely the result of the Arab slave trade from there. The second is that the eastward gene flow must have been fairly minor given the low level of haplotypes of Hgs L0-L6 shared between Ethiopians and Yemenis. Hg L6, first defined in this study, is the third passage but an intriguing one given its high frequency and no close match at that time.

The makeup of hgs of western Eurasian origin in Ethiopia is markedly different between the north-eastern and the south-central samples in the case of lineages belonging to macroHg N. The former is the Tigrinya region, where Hgs HV, TJ, U, N1 and W are more frequent. However, they have few exact matches with southern Arabia, a feature that is actually also found amongst the other Ethiopian samples. This shows that Semitic influences from southern Arabia is weakly supported by our data. Indeed, several of the N lineages found in Ethiopia are also found in Near East, the Caucasus and North Africa.

Overall, on a very general level, the Ethiopian and Yemeni gene pools may appear to reflect extensive gene flow between them given their very similar proportion of sub-Sharan and western Eurasian mtDNA hgs. Yet, when looking at lineages, a very different picture is drawn, showing that they were influenced by different sources.

After-publication service

Yemen has been studied in several publications. In their 2008 study, Cerny and his colleagues analysed Yemeni samples from 4 different locations (Cerny *et al.* 2008). They showed that the western samples had closer affinities with Middle Eastern and North African samples despite the geographical proximity to East Africa. It was the more eastern Hadramawt samples that showed a higher closeness to East Africa, with parts of their sub-Sharan fraction likely reflecting the Arabian slave trade. A subsequent study using an increased number of Yemeni samples and a wider comparative data set reached broadly similar results (Cerny *et al.* 2016). The study of Rowold et al. (Rowold *et al.* 2007) of Yemeni samples pointed to a limited gene flow, rather ancient, through the Horn of Africa. On the other hand, using Bayesian analyses, Vyas et al. (Vyas *et al.* 2016) have shown 4 migrations, with the first two, during the Holocene, involving the Horn of Africa at large. The first one is linked to the Arabian slave trade, a connection also revealed in the study of Fernandes et al. (Fernandes *et al.* 2015).

Ethiopia has also been a starting point for some studies. The analyses by Boattini et al. (Boattini *et al.* 2013) revealed a cluster that is widespread in Ethiopia, comprising other populations speaking Afro-Asiatic languages, with affinities with Yemen and Egypt, which they interpreted as evidence for Ethiopia to have acted a primary hub for recent human migrations. Using autosomal data, Pagani et al. (Pagani *et al.* 2012) showed a strong genetic structuring in East Africa and also that the non-African component found in Ethiopians likely represented gene flow that originated in the Levant, not in southern Arabia, around ≈ 3 KYA. However, in the study Hodgson et al. (Hodgson *et al.* 2014), this latter conclusion is contested, the authors arguing for the non-African component being much older.

5.2. Filling some gaps in the European mitochondrial landscape (Ref. II)

The then

We were interested in describing first the mtDNA gene pool of France in the European context and secondly, through a detailed and localised sampling based on historical regions, to see if there exists some sub-variation within France, with a focus on two specific regions. We genotyped in total 868 mtDNA genomes and combined them with previously published French samples.

A PCA plot using the overall frequencies of hgs found in France reveals that its variation places it close to its immediate neighbours (Figure S1). This essentially reflects the overall homogeneity of the European mtDNA gene pool when using a broad classification of hgs.

Within France itself, regions show some degrees of separation, with some outliers, specially Bearn and Provence, yet the latter position might be better explained by low sample size than by a real peculiar composition (Figure 2a, Table S1).

When focusing on two specific French regions, Bearn and Brittany, placed into the European context, some features were apparent.

In the case of the French Basques (the French Basque country overlaps with Bearn), on the PC plot (Figure 2b) it is interesting to see that in the first component, both French and Spanish Basques are driven towards an extreme, whilst in the second component, they are driven apart. The similarity lies in the high prevalence of hg H among both, whereas they are drastically dissimilar for the occurrence of the hgs U5/T (high among Spanish Basques) and hgs U4/J (high among French Basques). Noteworthy within hg J is the sub-clade J1c, of likely Near Eastern origin, which is particularly prevalent among French Basques. This, taken together with the fact that sub-clade T1, linked with the Neolithic diffusion, is only present among French Basques, questions the use of the whole Basques population as a representative set of the Palaeolithic European mtDNA gene pool (Dupanloup *et al.* 2004). Also, these discrepancies show that, although Basques share a common language, genetic drift acting on semi-isolated populations in valleys as in the case of Lapurdi, together with possible admixture from different sources, such factors were strong enough to leave a visible impact in the overall mtDNA diversity of the Basques.

Brittany shares obvious connections with the Celtic world, with various migration waves coming from the British Isles, but chiefly through its language. Breton belongs to the Brythonic branch, a branch also comprising Welsh. Genetically, some similarities are also present, for instance the high prevalence of hg I in Finistère and the British Isles, mirroring that of the cystic fibrosis mutation G551D (Cashman *et al.* 1995). Beside the British Isles, some connections also exist with Scandinavia in respect to hg U5a, hg very frequent in Scandinavia and in Finistère, but at this level of resolution it is hard to clearly resolve the directionality of its occurrence here; is it through Vikings invasion or is it a remnant of the colonization of northern Europe after the Last Glacial

Maximum. This connection is also revealed by the similar repartition of the chemokine receptor CCR5 variants (Libert *et al.* 1998).

After-publication service

Whilst France as a whole may not have received again some further attention (see below for an exception of sort), the same cannot be said of the Basques. To start on a specific note, and tempering our results regarding the J1c clade amongst French Basques, the study by Cardoso et al. (Cardoso *et al.* 2011) found a high prevalence of this sub-clade in Basques from Navarre. Broadening the focus, the study of Martínez-Cruz and colleagues (Martinez-Cruz *et al.* 2012) in particular sought to address some of what they considered short-comings from the multitudinous previous studies dealing with Basques' origins. Their conclusion was that the genetic structure found in present day Basques was mostly shaped during pre-Roman time, the result of tribal structuring connected to geography. They subsequently were less influenced by later gene flows that affected the surrounding regions, but yet still fall within the overall genetic Western European pool. That less pronounced influence could therefore make them potential representatives of older European populations. Several other studies support this notion; the one from Behar and colleagues (Behar *et al.* 2012a) showed a degree of genetic continuity with the preceding Paleolithic/ Mesolithic settlers, as the one from Carodoso et al. (Cardoso *et al.* 2013). Using ancient DNA, a similar conclusion was reached in Palencia-Madrid et al. (Palencia-Madrid *et al.* 2017), although they noted that "Iberians *[Basques]* might have been less affected by the Neolithic mitochondrial lineages".

Regarding Brittany, no further mtDNA study exist. Yet, the connections brought forth in Ref. II between Brittany and other Celtic regions have been further highlighted in different studies, but we will limit ourselves here to autosomal data. For instance, in the exception alluded above (Karakachoff *et al.* 2015), two cohorts of individuals from Western France were studied using the Axiom Genome-Wide CEU-1 Array. The authors showed a level of fine-scale population structure, as well as of low level of differentiation between Bretons and Irish at genome-side level and especially in the lactase region and HLA complex. In the study of Leslie et al. (Leslie *et al.* 2015), the group they named FRA14 comprises mostly samples from around the region of Rennes, an important city in Brittany. Pertinent here is the fact that this group "has its highest contributions in all western clusters (Cornwall, the three Welsh clusters, the cluster spanning Northern Ireland and western Scotland)", all regions of Celtic heritage.

5.3. Revisiting the North-African specific mtDNAs legacy (Ref. III)

The then

In this study, we refined the phylogenies of hgs M1 and U6 and used a corrected mutation rate in order to reassess the temporal aspect of their dispersal.

Several studies had focused on either or both; one crucial point in them was the claim of a southwest Asian origin for both hgs (Gonzalez *et al.* 2007; Maca-Meyer *et al.* 2003; Olivieri et al. 2006; Pereira *et al.* 2010). The updated phylogenetic trees and networks (Additional files 2-5 of Ref. III) drawn using our samples and the literature did not unequivocally support a southwestern Asian origin, that is M1 and U6 respective phylogeny are not rooted in a clade with distinct southwestern Asian origin. Further, none of their immediate sub-clades bifurcating from their root also shown a clear southwestern Asian tropism. This does not discard southwestern Asia as their place of origin; indeed, it still remains the most likely option, the simple fact is that there is no "hard" evidence for it, and it may remain forever elusive, unless some aDNA findings were to shed some light on this.

Another particular point was the timing (and to some extent the place) of M1 and U6 earliest expansions. As briefly discusses in section 2.2.2 and in Ref. III, the earlier studies relied on a method of molecular dating that was shown to be prone to over-estimation (Endicott et al. 2009; Loogvali et al. 2009; Soares et al. 2009), point partially addressed in the study by Pereira *et al.* (Pereira et al. 2010). Using the "Kivisild" mutation rate (Kivisild et al. 2006) refined in the study of Loogväli et al. (Loogvali et al. 2009), younger ages than previous studies were indeed obtained, with 32.8 ± 7.0 KYA for U6 and 28.9 ± 7.6 KYA for M1; ages, especially for M1, that cannot be easily reconciled with the Early Upper Palaeolithic expansion time put forth by Olivieri et al. (Olivieri et al. 2006).

This coalescent age discrepancy is only one of further evidences against a potential mimicry between expansion times of M1 and U6 that was put forth (Gonzalez et al. 2007; Olivieri et al. 2006; Pereira et al. 2010). The difference is also highlighted in the BSP (Drummond et al. 2005) analyses (Figure 3; Ref III) where it is visible that the mean curves of expansion times of the two hgs are hardly overlapping. Although both hgs seems to have experienced two more pronounced expansion phases, their intensity and timing do not overlap at all.

Yet, a common feature is that U6 and M1 mostly expanded within North (East) Africa. Few of their sub-clades can be tentatively associated with known archaeological industries. For instance, with a probable origin in North West Africa, together with the calculated coalescent ages around ≈20-22 KYA, U6a1 and M1b would slightly predate or coincide with the flourishing of the Iberomaurusian industry in the Maghreb (Barich and Garcea 2008; Barton *et al.* 2008; Merzoug and Sari 2008). In the same region, at around ≈10-11 KYA, U6b and M1b1 expansions coincide with the Caspian Industry (Barker *et al.* 2015; Barton et al. 2008; Garcea and Giraudi 2006).

Also, we investigated if there was a link between M1 and/or U6 dispersal(s) and that of the Afro-Asiatic languages. Such claims have been put forth earlier (Forster 2004; Forster and Romano 2007; Gonzalez et al. 2007; Maca-Meyer et al. 2003; Olivieri et al. 2006), but were never actually tested. Mantel tests show that for U6 together with M1; as a whole, or for U6 alone, no correlation exists between geography and language. However, such correlation does exist for M1 alone, and more particularly for M1a. Yet, if the genetic evidences may point towards a correlation between the spread of AA languages and M1a; hg with a likely East African origin, this does not discard the other linguistic hypothesis that posits an origin for AA languages in the Middle East.

After-publication service

Couple of years after the publication of Ref. III, Secher and colleagues (Secher *et al.* 2014) published a study focusing solely on U6, almost doubling the number of full sequences and considerably increasing the partially genotyped samples. This latter effort allowed them to turn their attention to some specific migrations events, such as U6 contribution to the American Continent. However, the major conclusions reached in their study largely confirmed the findings of Ref. III. The results of an ancient DNA study are also in line with our coalescent estimates (using a different mutation rate) and "asynchronous increase" in U6 and M1 N_e (van de Loosdrecht *et al.* 2018), whilst a second added further evidence to the prevalence of U6 and M1 amongst early inhabitants of North Africa (Fregel *et al.* 2018).

5.4. The Hindu Kush region of Afghanistan, vortex of various sources (Ref. IV)

The then

In this study, we present the results of more than 500 samples from 5 different ethnic origins (Pashtun, Tajik, Hazara, Uzbek and Turkmen) collected around the Hindu Kush region in Afghanistan. We characterised their mtDNA and chrY gene pools, and also genotyped a subset of 5 samples from each population (only 4 for Turkem) and put these results in the geographical context.

The autosomal analyses reveal a relative homogeneity between the different ethnic groups. The F_{st} distances between these groups, despite different linguistic affiliations, are rather small, even smaller than that with similar ethnic groups from neighbour countries (Figure S4 in Ref. IV). The PCA analysis (Figure 3 in Ref. IV) shows a slightly less compact cluster for the Hindu Kush population, while retaining a very strong geographic overlook, with Central Asian populations forming a loose cluster in the middle of the two components. Again, the East Asian component present in the Hazaras, likely brought during the Mongols invasion, is seen by their position closer the Mongols and East Asia as compared to their direct geographic neighbours. Congruent with the PCA, plotting several of the ancestry components (AC) resulting from the ADMIXTURE analysis

(Figure 2 in Ref. IV) exhibits the role of crossroad of Central Asia. Most notably, there is no AC that shows a clear Central Asian tropism, but rather many ACs converge towards Central Asia, or stop short of it. There is a rather strong sub-continental structuring in their repartition, which is also somewhat reminiscent of broad linguistic affiliations, a correlation lost at the level of the Hindu Kush populations. Indeed, looking at the more traditional bar plot way (Figures S1 and S2 in Ref. IV), the finer structure of the Hindu Kush populations can be seen, and there is no distinct difference explainable by linguistic amongst them. Their relative homogeneity, and that of Central Asia more generally, is once more plainly apparent. This homogeneity can tentatively be explained via the Inner Asian Mountain Corridor framework put forth by Frachetti (Frachetti 2012), where a common hunter-gatherer background sees extensive inter-regional pastoralism, with effect of geography playing a more major role than language, even in the light of major influx, like that of the Iranian origin supported by archaeological evidence (DQ 2009; Fuller 2007) and the Y chromosome hg J2a1-Page55 distribution (Table S6 in Ref. IV).

From a mitochondrial perspective, there are some differences between the different ethnic groups and even amongst themselves when compare to other populations with similar ethnic affiliation in other countries. One notable feature is the higher prevalence of typical East Eurasian mtDNA hgs among Hazaras, mirroring what can be seen from the chrY with C3b2b1-M401 (see below). When explored using a factorial analysis with several other populations, the global factors influencing the first two axes are the regional specific hgs, resulting in a triangular distribution. This distribution of the clusters reflects linguistic affiliations of the populations (Figure S6A-C in Ref. IV), clusters which are less pronounced when looking at the geographical origin, especially in the case of Central Asia.

Regarding the chrY, we refined the topology of hg C3-PK2, linked with the Mongol invasion (Zerjal *et al.* 2003). Two new SNPs allow to divide C3-PK2 into C3a-M386 and C3b-M532; into which all our C3-PK2 samples fall. C3b-M532 is also further divided downstream, most notably C3b2b1-M401 that encompasses the Mongol 'star cluster' YSTR haplotype and is found in Afghan Hazara, Mongols as well as in Kyrgyz. This refined phylogeny should help narrowing down the descendants of the Mongol invasions.

What transpires from the different analyses is that, notwithstanding the clear diverse linguistic and ethnic diversity of the Hindu Kush populations studied, their underlying gene pools are broadly similar. They reflect gene flow from West and East Eurasia, as well as from South Asia, through post-glacial expansions; for instance, stemming from the Fertile Crescent in the early Neolithic times. Later, more recent layers, have also left an imprint as seen with the "historical" example of the Hazaras.

After-publication service

A study published in 2018 (Peng *et al.* 2018) presented samples from the Pamir populations, located north-east of the Hindu Kush range. The authors also did

not find any basal lineages in Central Asia, further reinforcing the findings that Central Asians are the results of "an East-West Eurasian admixture". The study of Jeong and colleagues (Jeong *et al.* 2019) has a much broader geographical focus, yet some of these results, particularly Figure 4 and the conspicuous absence of a Central Asian group in the GLOBETROTTER, are also in line with our main conclusions. The connection between Central Asia and the Mongol empire has also been addressed several times in subsequent studies, amongst others (Hellenthal et al. 2014; Yunusbayev *et al.* 2015).

6. CONCLUSION

If one were inclined to forgo reading the previous chapter, below are the main conclusions:

REF I

- The global proportion of Eurasian and African specific lineages is almost identical in Ethiopian and Yemeni mtDNA gene pool.
- The study of haplotypes reveals finer details, such as lineages in South Arabia linked to East Africa and the higher abundance of N lineages of western Eurasian origin in northeast Ethiopia.

REF II

- The overall mtDNA variation of France meshes in the European context.
- The Basques north of the Pyrénées exhibit genetic features differentiating them from Basque on the south side.
- Finistère displays tighter connections with Britain and Scandinavia than the other administrative departments of Brittany.

REF III

- The respective phylogenies of haplogroups M1 and U6 do not support a clear southwestern Asian origin.
- The expansion times and putative locations of occurrence of their sub-clades largely do not overlap.
- No strong correlations between the spread of Afro-Asiatic languages and both haplogroups were found.

REF IV

- The Afghan Hindu Kush reflects the notion that Central Asia has been a long-standing cross-road of multiple waves of migrations, each leaving perceptible traces in nowadays different genomes of various ethnic populations.

SUMMARY IN ESTONIAN

Uidates mtDNA fülogeneesi radadel; essee väikeste kõrvalepõigetega sellest, mida see meile inimese migratsioonidest kõnelda võib

(Mõned) Inimesed on loomu poolest väga uudishimulikud ning pole tavatu, et neid huvitavad küsimused sellest, kust me tuleme. Neile küsimustele saab vastuseid otsida populatsioonigeneetika raamistikku kasutades. Populatsioonide ja nende päritolu kohta tekkinud küsimustele vastamiseks on populatsioonigeneetikal kasutada tõhus ja püsivalt arenev teooriate ja meetodite mahukas pakett.

Põhiline algmaterjal vastuste otsimisel – nagu populatsioonigeneetika nimestki tuletub – on genoom. Lõviosa inimgenoomist asub rakutuumas, kust leiame 23 kromosoomipaari. Üks paaridest on sugukromosoomid, neid nimetatakse X ja Y kromosoomideks, teised aga autosoomid. Y kromosoom on ainult meestel, naistel on kaks X kromosoomi. Ülejäänud väike osa genoomist asub rakuorganellis – mitokondris – ja seda nimetatakse mitokondriaalseks DNAks (mtDNA). Erinevate genoomi osade pärandumise täiesti erinev laad mängib nende mitmekesisuse tõlgendamisel otsustavat rolli. Y kromosoom ja mtDNA päranduvad vanemalt järglasele haploidsena, vaid isalt või emalt. Veelgi enam – Y kromosoom ja mtDNA, erinevalt ülejäänud genoomist, ei rekombineeru. See omadus teeb Y kromosoomi ja mtDNA fülogeneesi rekonstrueerimiseks äärmiselt sobilikuks ja autosoomidega võrreldes suhteliselt lihtsaks. Lisaväärtuseks on, et saadud fülogeneesipuud peegeldavad spetsiifiliselt meie ema- ja isaliinide genealoogiat.

Fülogeneesipuu rekonstrueerimine on ainult esimene samm, seda täiendavad teised analüüsid. Lähtudes molekulaarse kella kontseptsioonist on meil võimalik fülogeneesipuu erinevate harude, millele sageli viidatakse kui haplogruppidele, ligikaudne ekspansiooniaeg välja arvutada. Kui kombineerime haplogruppide vanuse nende geograafilise levikumustriga (fülogeograafia), saame fülogeneesipuule lisada ajalis-ruumilise mõõtme. Populatsioonide migratsioone puudutavatele küsimustele vastuste otsimisel aitab meid geneetiliste andmete kõrval muu teave, millest enim kasutatakse paleoantropoloogia, arheoloogia, ajaloo ja keeleteaduse teadmisi. Tänu neile saame me seniseid hüpoteese täiendada või isegi uusi püstitada ja neid siis geneetiliste andmete valguses testida.

Kui kõik nimetatu on arvesse võetud, peaksime olema võimelised ütlema, kust meie poolt uuritavad populatsioonid tulnud on ning meil tekivad head võimalused ka vaadeldava mitmekesisuse põhjuste lahti seletamiseks. Käesolev doktoritöö koosneb neljast uurimistööst, mis tuginevad ülalpool väga põgusalt kirjeldatud kontseptsioonile ja mille põhilised tulemused näitavad alljärgnevat:

REF I

- Euraasiale ja Aafrikale iseloomulike liinide üldine osakaal on Etioopia ja Jeemeni mtDNA geenitiigis identne.
- Haplotüüpide uurimine toob esile täpsemad detailid, näiteks on Lõuna-Araabia liinid seotud Ida-Aafrikaga ja Lääne-Euraasia päritolu N-liinide rohkus omane Kirde-Etioopiale.

REF II

- Prantsusmaa üldine mtDNA mitmekesisus sobitub Euroopa konteksti.
- Püreneedest põhjas elavatel baskidel on geneetilisi tunnuseid, mis eristavad neid lõunas elavatest baskides
- Finistère`l on Bretagne`i teiste administratiivsete piirkondadega võrreldes Suurbritannia ja Skandinaaviaga tugevamad sidemed.

REF III

- Haplogruppide M1 ja U6 fülogenees ei toeta hüpoteesi nende selge Edela-Aasia päritolu kohta.
- Nende alamklaadide ekspansiooniajad ja arvatavad levikukeskmed ei kattu.
- Nende haplogruppide ja afroaasia keelte leviku vahel ei ole tugevat korrelatsiooni.

REFERENCES

Alexander, D. H., J. Novembre and K. Lange. 2009. “Fast Model-Based Estimation of Ancestry in Unrelated Individuals.” *Genome Res* 19(9):1655–64.

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. de Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. Smith, R. Staden and I. G. Young. 1981. “Sequence and Organization of the Human Mitochondrial Genome.” *Nature* 290(5806):457–65.

Andrews, R. M., I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull and N. Howell. 1999. “Reanalysis and Revision of the Cambridge Reference Sequence for Human Mitochondrial DNA.” *Nat Genet* 23(2):147.

Avise, J., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Rub and N. C. Saunders. 1987. “Intraspecific Phylogeography: The Molecular Bridge between Population Genetics and Systematics.” *Ann. Rev. Ecol. Syst.* 18:489–522.

Avise, J. C. 2000. *Phylogeography: The History and Formation of Species.* Cambridge, Massachusetts: Harvard University Press.

Balanovsky, O. 2017. “Toward a Consensus on Snp and Str Mutation Rates on the Human Y-Chromosome.” *Hum Genet* 136(5):575–90.

Bandelt, H.-J., P. Forster, B. C. Sykes and M. B. Richards. 1995. “Mitochondrial Portraits of Human Populations Using Median Networks.” *Genetics* 141(2):743–53.

Bandelt, H.-J., P. Forster and A. Röhl. 1999. “Median-Joining Networks for Inferring Intraspecific Phylogenies.” *Mol Biol Evol* 16(1):37–48.

Barich, Barbara E. and Elena A. A. Garcea. 2008. “Ecological Patterns in the Upper Pleistocene and Holocene in the Jebel Gharbi, Northern Libya: Chronology, Climate and Human Occupation.” *African Archaeological Review* 25(1):87–97.

Barker, Graeme, Annita Antoniadou, Simon Armitage, Ian Brooks, Ian Candy, Kate Connell, Katerina Douka, Nicholas Drake, Lucy Farr, Evan Hill, Chris Hunt, Robyn Inglis, Sacha Jones, Christine Lane, Giulio Lucarini, John Meneely, Jacob Morales, Giuseppina Mutri, Amy Prendergast, Ryan Rabett, Hazel Reade, Tim Reynolds, Natalie Russell, David Simpson, Bernard Smith, Chris Stimpson, Mohammed Twati and Kevin White. 2015. “The Cyrenaican Prehistory Project 2010: The Fourth Season of Investigations of the Haua Fteah Cave and Its Landscape, and Further Results from the 2007–2009 Fieldwork.” *Libyan Studies* 41:63–88.

Barton, Nick, Abdeljalil Bouzouggar, Louise Humphrey, Peter Berridge, Simon Collcutt, Rowena Gale, Simon Parfitt, Adrian Parker, Edward Rhodes and Jean-Luc Schwenninger. 2008. “Human Burial Evidence from Hattab Ii Cave and the Question of Continuity in Late Pleistocene–Holocene Mortuary Practices in Northwest Africa.” *Cambridge Archaeological Journal* 18(2):195–214.

Behar, D. M., B. Yunusbayev, M. Metspalu, E. Metspalu, S. Rosset, J. Parik, S. Rootsi, G. Chaubey, I. Kutuev, G. Yudkovsky, E. K. Khusnutdinova, O. Balanovsky, O. Semino, L. Pereira, D. Comas, D. Gurwitz, B. Bonne-Tamir, T. Parfitt, M. F. Hammer, K. Skorecki and R. Villems. 2010. “The Genome-Wide Structure of the Jewish People.” *Nature* 466(7303):238–42.

Behar, D. M., C. Harmant, J. Manry, M. van Oven, W. Haak, B. Martinez-Cruz, J. Salaberria, B. Oyharcabal, F. Bauduer, D. Comas, L. Quintana-Murci and Consortium Genographic. 2012a. “The Basque Paradigm: Genetic Evidence of a Maternal Continuity in the Franco-Cantabrian Region since Pre-Neolithic Times.” *Am J Hum Genet* 90(3):486–93.

Behar, D. M., M. van Oven, S. Rosset, M. Metspalu, E. L. Loogvali, N. M. Silva, T. Kivisild, A. Torroni and R. Villems. 2012b. "A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from Its Root." *Am J Hum Genet* 90(4):675–84.

Behar, Doron M., Richard Villems, Himla Soodyall, Jason Blue-Smith, Luisa Pereira, Ene Metspalu, Rosaria Scozzari, Heeran Makkan, Shay Tzur, David Comas, Jaume Bertranpetit, Lluis Quintana-Murci, Chris Tyler-Smith, R. Spencer Wells and Saharon Rosset. 2008. "The Dawn of Human Matrilineal Diversity." *The American Journal of Human Genetics* 82(5):1130–40.

Bellott, D. W., J. F. Hughes, H. Skaletsky, L. G. Brown, T. Pyntikova, T. J. Cho, N. Koutseva, S. Zaghlul, T. Graves, S. Rock, C. Kremitzki, R. S. Fulton, S. Dugan, Y. Ding, D. Morton, Z. Khan, L. Lewis, C. Buhay, Q. Wang, J. Watt, M. Holder, S. Lee, L. Nazareth, J. Alfoldi, S. Rozen, D. M. Muzny, W. C. Warren, R. A. Gibbs, R. K. Wilson and D. C. Page. 2014. "Mammalian Y Chromosomes Retain Widely Expressed Dosage-Sensitive Regulators." *Nature* 508(7497):494–9.

Boattini, A., L. Castri, S. Sarno, A. Useli, M. Cioffi, M. Sazzini, P. Garagnani, S. De Fanti, D. Pettener and D. Luiselli. 2013. "Mtdna Variation in East Africa Unravels the History of Afro-Asiatic Groups." *Am J Phys Anthropol* 150(3):375–85.

Bromham, L. and D. Penny. 2003. "The Modern Molecular Clock." *Nat Rev Genet* 4(3):216–24.

Brotherton, P., W. Haak, J. Templeton, G. Brandt, J. Soubrier, C. Jane Adler, S. M. Richards, C. Der Sarkissian, R. Ganslmeier, S. Friederich, V. Dresely, M. van Oven, R. Kenyon, M. B. Van der Hoek, J. Korlach, K. Luong, S. Y. W. Ho, L. Quintana-Murci, D. M. Behar, H. Meller, K. W. Alt, A. Cooper and Consortium Genographic. 2013. "Neolithic Mitochondrial Haplogroup H Genomes and the Genetic Origins of Europeans." *Nat Commun* 4:1764.

Brown, W. M. 1980. "Polymorphism in Mitochondrial DNA of Humans as Revealed by Restriction Endonuclease Analysis." *Proc Natl Acad Sci U S A* 77(6):3605–9.

Cann, R. L., M. Stoneking and A. C. Wilson. 1987. "Mitochondrial DNA and Human Evolution." *Nature* 325(6099):31–6.

Cardoso, S., M. A. Alfonso-Sanchez, L. Valverde, A. Odriozola, A. M. Perez-Miranda, J. A. Pena and M. M. de Pancorbo. 2011. "The Maternal Legacy of Basques in Northern Navarre: New Insights into the Mitochondrial DNA Diversity of the Franco-Cantabrian Area." *Am J Phys Anthropol* 145(3):480–8.

Cardoso, S., L. Valverde, M. A. Alfonso-Sanchez, L. Palencia-Madrid, X. Elcoro-aristizabal, J. Algorta, S. Catarino, D. Arteta, R. J. Herrera, M. T. Zarrabeitia, J. A. Pena and M. M. de Pancorbo. 2013. "The Expanded Mtdna Phylogeny of the Franco-Cantabrian Region Upholds the Pre-Neolithic Genetic Substrate of Basques." *PLoS One* 8(7):e67835.

Carelli, V. 2015. "Keeping in Shape the Dogma of Mitochondrial DNA Maternal Inheritance." *PLoS Genet* 11(5):e1005179.

Cashman, S. M., A. Patino, A. Martinez, M. Garcia-Delgado, Z. Miedzybrodzka, M. Schwarz, A. Shrimpton, C. Ferec, O. Raguenes, M. Macek, Jr. and et al. 1995. "Identical Intragenic Microsatellite Haplotype Found in Cystic Fibrosis Chromosomes Bearing Mutation G551d in Irish, English, Scottish, Breton and Czech Patients." *Hum Hered* 45(1):6–12.

Cavalli-Sforza, L. L., A. C. Wilson, C. R. Cantor, R. M. Cook-Deegan and M. C. King. 1991. "Call for a Worldwide Survey of Human Genetic Diversity: A Vanishing Opportunity for the Human Genome Project." *Genomics* 11(2):490–1.

Cavalli-Sforza, L. L., Paolo Menozzi and Alberto Piazza. 1994. *The History and Geography of Human Genes*. Princeton, N.J.: Princeton University Press.

Cerny, V., C. J. Mulligan, J. Ridl, M. Zaloudkova, C. M. Edens, M. Hajek and L. Pereira. 2008. "Regional Differences in the Distribution of the Sub-Saharan, West Eurasian, and South Asian Mtdna Lineages in Yemen." *Am J Phys Anthropol* 136(2):128–37.

Cerny, V., M. Cizkova, E. S. Poloni, A. Al-Meeri and C. J. Mulligan. 2016. "Comprehensive View of the Population History of Arabia as Inferred by Mtdna Variation." *Am J Phys Anthropol* 159(4):607–16.

Charlesworth, B. 2009. "Fundamental Concepts in Genetics: Effective Population Size and Patterns of Molecular Evolution and Variation." *Nat Rev Genet* 10(3):195–205.

Chaubey, G., Q. Ayub, N. Rai, S. Prakash, V. Mushrif-Tripathy, M. Mezzavilla, A. K. Pathak, R. Tamang, S. Firasat, M. Reidla, M. Karmin, D. S. Rani, A. G. Reddy, J. Parik, E. Metspalu, S. Rootsi, K. Dalal, S. Khaliq, S. Q. Mehdi, L. Singh, M. Metspalu, T. Kivisild, C. Tyler-Smith, R. Villems and K. Thangaraj. 2017. ""Like Sugar in Milk": Reconstructing the Genetic History of the Parsi Population." *Genome Biol* 18(1):110.

Chinnery, P. F. 2006. "Mitochondrial DNA in Homo Sapiens." Pp. pp. 3–15 in *Human Mitochondrial DNA and Hte Evolution of Homo Sapiens*, edited by H. J. M. Bandelt, V.; Richards, M.: Springer-Verlag Berlin Heidelberg.

Cortez, D., R. Marin, D. Toledo-Flores, L. Froidevaux, A. Liechti, P. D. Waters, F. Grutzner and H. Kaessmann. 2014. "Origins and Functional Evolution of Y Chromosomes across Mammals." *Nature* 508(7497):488–93.

Cox, M. P. 2008. "Accuracy of Molecular Dating with the Rho Statistic: Deviations from Coalescent Expectations under a Range of Demographic Models." *Hum Biol* 80(4):335–57.

Denaro, M., H. Blanc, M. J. Johnson, K. H. Chen, E. Wilmsen, L. L. Cavalli-Sforza and D. C. Wallace. 1981. "Ethnic Variation in Hpa 1 Endonuclease Cleavage Patterns of Human Mitochondrial DNA." *Proc Natl Acad Sci U S A* 78(9):5768–72.

DQ, Fuller. 2009. "Framing a Middle Asian Corridor of Crops Exchange and Agricultural Innovation." Pp. p 3 in *13th Harvard University Round Table. Ethnogenesis of South and Central Asia*. Kyoto: Research Insitute for Humanity and Nature.

Drummond, A. J., A. Rambaut, B. Shapiro and O. G. Pybus. 2005. "Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences." *Mol Biol Evol* 22(5):1185–92.

Drummond, A. J. and A. Rambaut. 2007. "Beast: Bayesian Evolutionary Analysis by Sampling Trees." *BMC Evol Biol* 7:214.

Drummond, A. J., M. A. Suchard, D. Xie and A. Rambaut. 2012. "Bayesian Phylogenetics with Beauti and the Beast 1.7." *Mol Biol Evol* 29(8):1969–73.

Duan, M., J. Tu and Z. Lu. 2018. "Recent Advances in Detecting Mitochondrial DNA Heteroplasmic Variations." *Molecules* 23(2).

Dupanloup, I., G. Bertorelle, L. Chikhi and G. Barbujani. 2004. "Estimating the Impact of Prehistoric Admixture on the Genome of Europeans." *Mol Biol Evol* 21(7):1361–72.

Elson, J. L., D. M. Turnbull and N. Howell. 2004. "Comparative Genomics and the Evolution of Human Mitochondrial DNA: Assessing the Effects of Selection." *Am J Hum Genet* 74(2):229–38.

Endicott, P. and S. Y. Ho. 2008. “A Bayesian Evaluation of Human Mitochondrial Substitution Rates.” *Am J Hum Genet* 82(4):895–902.

Endicott, P., S. Y. Ho, M. Metspalu and C. Stringer. 2009. “Evaluating the Mitochondrial Timescale of Human Evolution.” *Trends Ecol Evol* 24(9):515–21.

Evanno, G., S. Regnaut and J. Goudet. 2005. “Detecting the Number of Clusters of Individuals Using the Software Structure: A Simulation Study.” *Mol Ecol* 14(8): 2611–20.

Falush, D.; van Dorp, L.; Lawson D:. 2016. “A Tutorial on How (Not) to over-Interpret Structure/Admixture Bar Plots.” *bioRxiv 066431*.

Fernandes, V., P. Triska, J. B. Pereira, F. Alshamali, T. Rito, A. Machado, Z. Fajkosova, B. Cavadas, V. Cerny, P. Soares, M. B. Richards and L. Pereira. 2015. “Genetic Stratigraphy of Key Demographic Events in Arabia.” *PLoS One* 10(3):e0118625.

Filosto, M., M. Mancuso, C. Vives-Bauza, M. R. Vila, S. Shanske, M. Hirano, A. L. Andreu and S. DiMauro. 2003. “Lack of Paternal Inheritance of Muscle Mitochondrial DNA in Sporadic Mitochondrial Myopathies.” *Ann Neurol* 54(4):524–6.

Foote, S., D. Vollrath, A. Hilton and D. C. Page. 1992. “The Human Y Chromosome: Overlapping DNA Clones Spanning the Euchromatic Region.” *Science* 258(5079): 60–6.

Forster, P., R. Harding, A. Torroni and H.-J. Bandelt. 1996. “Origin and Evolution of Native American Mtdna Variation: A Reappraisal.” *Am J Hum Genet* 59(4):935–45.

Forster, P. 2004. “Ice Ages and the Mitochondrial DNA Chronology of Human Dispersals: A Review.". *Philos Trans R Soc Lond B Biol Sci.* 359(1442):255–64.

Forster, P. and V. Romano. 2007. “Timing of a Back-Migration into Africa.” *Science* 316(5821):50–3.

Frachetti, M. D. 2012. “Multiregional Emergence of Mobile Pastoralism and Nonuniform Institutional Complexity across Eurasia.” *Current Anthropology* 53(1):2–38.

Francois, O., M. Currat, N. Ray, E. Han, L. Excoffier and J. Novembre. 2010. “Principal Component Analysis under Population Genetic Models of Range Expansion and Admixture.” *Mol Biol Evol* 27(6):1257–68.

Fregel, Rosa, Fernando L. Méndez, Youssef Bokbot, Dimas Martín-Socas, María D. Camalich-Massieu, Jonathan Santana, Jacob Morales, María C. Ávila-Arcos, Peter A. Underhill, Beth Shapiro, Genevieve Wojcik, Morten Rasmussen, André E. R. Soares, Joshua Kapp, Alexandra Sockell, Francisco J. Rodríguez-Santos, Abdeslam Mikdad, Aioze Trujillo-Mederos and Carlos D. Bustamante. 2018. “Ancient Genomes from North Africa Evidence Prehistoric Migrations to the Maghreb from Both the Levant and Europe.” *Proceedings of the National Academy of Sciences* 115(26):6774.

Fu, Q., A. Mittnik, P. L. F. Johnson, K. Bos, M. Lari, R. Bollongino, C. Sun, L. Giemsch, R. Schmitz, J. Burger, A. M. Ronchitelli, F. Martini, R. G. Cremonesi, J. Svoboda, P. Bauer, D. Caramelli, S. Castellano, D. Reich, S. Paabo and J. Krause. 2013. “A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes.” *Curr Biol* 23(7):553–59.

Fuller, Dorian Q. 2007. “Contrasting Patterns in Crop Domestication and Domestication Rates: Recent Archaeobotanical Insights from the Old World.” *Annals of Botany* 100(5):903–24.

Garcea, E. A. and C. Giraudi. 2006. “Late Quaternary Human Settlement Patterning in the Jebel Gharbi.” *J Hum Evol* 51(4):411–21.

Genomes Project, Consortium, G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles and G. A. McVean. 2010. “A Map of Human Genome Variation from Population-Scale Sequencing.” *Nature* 467(7319):1061–73.

Genomes Project, Consortium, G. R. Abecasis, A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth and G. A. McVean. 2012. “An Integrated Map of Genetic Variation from 1,092 Human Genomes.” *Nature* 491(7422):56–65.

Genomes Project, Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean and G. R. Abecasis. 2015. “A Global Reference for Human Genetic Variation.” *Nature* 526(7571):68–74.

Gibbons, A. 2011. “Anthropology. A New View of the Birth of Homo Sapiens.” *Science* 331(6016):392–4.

Giles, R. E., H. Blanc, H. M. Cann and D. C. Wallace. 1980. “Maternal Inheritance of Human Mitochondrial DNA.” *Proc Natl Acad Sci U S A* 77(11):6715–9.

Gonzalez, A. M., J. M. Larruga, K. K. Abu-Amero, Y. Shi, J. Pestano and V. M. Cabrera. 2007. “Mitochondrial Lineage M1 Traces an Early Human Backflow to Africa.” *BMC Genomics* 8:223.

Graves, J. A. 1995. “The Evolution of Mammalian Sex Chromosomes and the Origin of Sex Determining Genes.” *Philos Trans R Soc Lond B Biol Sci* 350(1333):305–11; discussion 11–2.

Green, R. E., J. Krause, S. E. Ptak, A. W. Briggs, M. T. Ronan, J. F. Simons, L. Du, M. Egholm, J. M. Rothberg, M. Paunovic and S. Paabo. 2006. “Analysis of One Million Base Pairs of Neanderthal DNA.” *Nature* 444(7117):330–6.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prufer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Hober, B. Hoffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich and S. Paabo. 2010. “A Draft Sequence of the Neandertal Genome.” *Science* 328(5979):710–22.

Haak, W., P. Forster, B. Bramanti, S. Matsumura, G. Brandt, M. Tanzer, R. Villems, C. Renfrew, D. Gronenborn, K. W. Alt and J. Burger. 2005. “Ancient DNA from the First European Farmers in 7500-Year-Old Neolithic Sites.” *Science* 310(5750): 1016–18.

Haak, W., O. Balanovsky, J. J. Sanchez, S. Koshel, V. Zaporozhchenko, C. J. Adler, C. S. Der Sarkissian, G. Brandt, C. Schwarz, N. Nicklisch, V. Dresely, B. Fritsch, E. Balanovska, R. Villems, H. Meller, K. W. Alt, A. Cooper and Consortium Members of the Genographic. 2010. “Ancient DNA from European Early Neolithic Farmers Reveals Their near Eastern Affinities.” *PLoS Biol* 8(11):e1000536.

Helgason, A., A. W. Einarsson, V. B. Guethmundsdottir, A. Sigurethsson, E. D. Gunnarsdottir, A. Jagadeesan, S. S. Ebenesersdottir, A. Kong and K. Stefansson. 2015. “The Y-Chromosome Point Mutation Rate in Humans.” *Nat Genet* 47(5):453–7.

Hellenthal, G., G. B. J. Busby, G. Band, J. F. Wilson, C. Capelli, D. Falush and S. Myers. 2014. “A Genetic Atlas of Human Admixture History.” *Science* 343(6172): 747–51.

Heller, R., L. Chikhi and H. R. Siegismund. 2013. “The Confounding Effect of Population Structure on Bayesian Skyline Plot Inferences of Demographic History.” *PLoS One* 8(5):e62992.

Henn, B. M., C. R. Gignoux, M. W. Feldman and J. L. Mountain. 2009. “Characterizing the Time Dependency of Human Mitochondrial DNA Mutation Rate Estimates.” *Mol Biol Evol* 26(1):217–30.

Heyer, E., J. Puymirat, P. Dieltjes, E. Bakker and P. de Knijff. 1997. “Estimating Y Chromosome Specific Microsatellite Mutation Frequencies Using Deep Rooting Pedigrees.” *Hum Mol Genet* 6(5):799–803.

Heyer, E., E. Zietkiewicz, A. Rochowski, V. Yotova, J. Puymirat and D. Labuda. 2001. “Phylogenetic and Familial Estimates of Mitochondrial Substitution Rates: Study of Control Region Mutations in Deep-Rooting Pedigrees.” *Am J Hum Genet* 69(5): 1113–26.

Hodgson, J. A., C. J. Mulligan, A. Al-Meeri and R. L. Raaum. 2014. “Early Back-to-Africa Migration into the Horn of Africa.” *PLoS Genet* 10(6):e1004393.

Holt, I. J., J. He, C. C. Mao, J. D. Boyd-Kirkup, P. Martinsson, H. Sembongi, A. Reyes and J. N. Spelbrink. 2007. “Mammalian Mitochondrial Nucleoids: Organizing an Independently Minded Genome.” *Mitochondrion* 7(5):311–21.

Howell, N., C. B. Smejkal, D. A. Mackey, P. F. Chinnery, D. M. Turnbull and C. Herrnstadt. 2003. “The Pedigree Rate of Sequence Divergence in the Human Mitochondrial Genome: There Is a Difference between Phylogenetic and Pedigree Rates.” *Am J Hum Genet* 72(3):659–70.

Howells, W. W. 1976. “Explaining Modern Man: Evolutionists Versus Migrationists.” *Journal of Human Evolution* 5(5):pp. 477–95.

Hudson, R. R. 1983. “Properties of a Neutral Allele Model with Intragenic Recombination.” *Theor Popul Biol* 23(2):183–201.

Huson, D.H., R. Rupp and C. Scornavacca. 2011. *Phylogenetic Networks - Concepts, Algorithms and Application*. Cambridge: Cambridge University Press.

Ingman, M., H. Kaessmann, S. Pääbo and U. Gyllensten. 2000. “Mitochondrial Genome Variation and the Origin of Modern Humans.” *Nature* 408:708–13.

Ingman, M. and U. Gyllensten. 2001. “Analysis of the Complete Human Mtdna Genome: Methodology and Inferences for Human Evolution.” *J Hered* 92(6):454–61.

International HapMap, Consortium. 2005. “A Haplotype Map of the Human Genome.” *Nature* 437(7063):1299–320.

International HapMap, Consortium, K. A. Frazer, D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve, R. A. Gibbs, J. W. Belmont, A. Boudreau, P. Hardenbol, S. M. Leal, S. Pasternak, D. A. Wheeler, T. D. Willis, F. Yu, H. Yang, C. Zeng, Y. Gao, H. Hu, W. Hu, C. Li, W. Lin, S. Liu, H. Pan, X. Tang, J. Wang, W. Wang, J. Yu, B. Zhang, Q. Zhang, H. Zhao, H. Zhao, J. Zhou, S. B. Gabriel, R. Barry, B. Blumenstiel, A. Camargo, M. Defelice, M. Faggart, M. Goyette, S. Gupta, J. Moore, H. Nguyen, R. C. Onofrio, M. Parkin, J. Roy, E. Stahl, E. Winchester, L. Ziaugra, D. Altshuler, Y. Shen, Z. Yao, W. Huang, X. Chu, Y. He, L. Jin, Y. Liu, Y. Shen, W. Sun, H. Wang, Y. Wang, Y. Wang, X. Xiong, L. Xu, M. M. Waye, S. K. Tsui, H. Xue, J. T. Wong, L. M. Galver, J. B. Fan, K. Gunderson, S. S. Murray, A. R. Oliphant, M. S. Chee, A. Montpetit, F. Chagnon, V. Ferretti, M. Leboeuf, J. F.

Olivier, M. S. Phillips, S. Roumy, C. Sallee, A. Verner, T. J. Hudson, P. Y. Kwok, D. Cai, D. C. Koboldt, R. D. Miller, L. Pawlikowska, P. Taillon-Miller, M. Xiao, L. C. Tsui, W. Mak, Y. Q. Song, P. K. Tam, Y. Nakamura, T. Kawaguchi, T. Kitamoto, T. Morizono, A. Nagashima, Y. Ohnishi, A. Sekine, T. Tanaka, T. Tsunoda, P. Deloukas, C. P. Bird, M. Delgado, E. T. Dermitzakis, R. Gwilliam, S. Hunt, J. Morrison, D. Powell, B. E. Stranger, P. Whittaker, D. R. Bentley, M. J. Daly, P. I. de Bakker, J. Barrett, Y. R. Chretien, J. Maller, S. McCarroll, N. Patterson, I. Pe'er, A. Price, S. Purcell, D. J. Richter, P. Sabeti, R. Saxena, S. F. Schaffner, P. C. Sham, P. Varilly, D. Altshuler, L. D. Stein, L. Krishnan, A. V. Smith, M. K. Tello-Ruiz, G. A. Thorisson, A. Chakravarti, P. E. Chen, D. J. Cutler, C. S. Kashuk, S. Lin, G. R. Abecasis, W. Guan, Y. Li, H. M. Munro, Z. S. Qin, D. J. Thomas, G. McVean, A. Auton, L. Bottolo, N. Cardin, S. Eyheramendy, C. Freeman, J. Marchini, S. Myers, C. Spencer, M. Stephens, P. Donnelly, L. R. Cardon, G. Clarke, D. M. Evans, A. P. Morris, B. S. Weir, T. Tsunoda, J. C. Mullikin, S. T. Sherry, M. Feolo, A. Skol, H. Zhang, C. Zeng, H. Zhao, I. Matsuda, Y. Fukushima, D. R. Macer, E. Suda, C. N. Rotimi, C. A. Adebamowo, I. Ajayi, T. Aniagwu, P. A. Marshall, C. Nkwodimmah, C. D. Royal, M. F. Leppert, M. Dixon, A. Peiffer, R. Qiu, A. Kent, K. Kato, N. Niikawa, I. F. Adewole, B. M. Knoppers, M. W. Foster, E. W. Clayton, J. Watkin, R. A. Gibbs, J. W. Belmont, D. Muzny, L. Nazareth, E. Sodergren, G. M. Weinstock, D. A. Wheeler, I. Yakub, S. B. Gabriel, R. C. Onofrio, D. J. Richter, L. Ziaugra, B. W. Birren, M. J. Daly, D. Altshuler, R. K. Wilson, L. L. Fulton, J. Rogers, J. Burton, N. P. Carter, C. M. Clee, M. Griffiths, M. C. Jones, K. McLay, R. W. Plumb, M. T. Ross, S. K. Sims, D. L. Willey, Z. Chen, H. Han, L. Kang, M. Godbout, J. C. Wallenburg, P. L'Archeveque, G. Bellemare, K. Saeki, H. Wang, D. An, H. Fu, Q. Li, Z. Wang, R. Wang, A. L. Holden, L. D. Brooks, J. E. McEwen, M. S. Guyer, V. O. Wang, J. L. Peterson, M. Shi, J. Spiegel, L. M. Sung, L. F. Zacharia, F. S. Collins, K. Kennedy, R. Jamieson and J. Stewart. 2007. “A Second Generation Human Haplotype Map of over 3.1 Million Snps.” *Nature* 449(7164):851–61.

International HapMap, Consortium, D. M. Altshuler, R. A. Gibbs, L. Peltonen, D. M. Altshuler, R. A. Gibbs, L. Peltonen, E. Dermitzakis, S. F. Schaffner, F. Yu, L. Peltonen, E. Dermitzakis, P. E. Bonnen, D. M. Altshuler, R. A. Gibbs, P. I. de Bakker, P. Deloukas, S. B. Gabriel, R. Gwilliam, S. Hunt, M. Inouye, X. Jia, A. Palotie, M. Parkin, P. Whittaker, F. Yu, K. Chang, A. Hawes, L. R. Lewis, Y. Ren, D. Wheeler, R. A. Gibbs, D. M. Muzny, C. Barnes, K. Darvishi, M. Hurles, J. M. Korn, K. Kristiansson, C. Lee, S. A. McCarrol, J. Nemesh, E. Dermitzakis, A. Keinan, S. B. Montgomery, S. Pollack, A. L. Price, N. Soranzo, P. E. Bonnen, R. A. Gibbs, C. Gonzaga-Jauregui, A. Keinan, A. L. Price, F. Yu, V. Anttila, W. Brodeur, M. J. Daly, S. Leslie, G. McVean, L. Moutsianas, H. Nguyen, S. F. Schaffner, Q. Zhang, M. J. Ghori, R. McGinnis, W. McLaren, S. Pollack, A. L. Price, S. F. Schaffner, F. Takeuchi, S. R. Grossman, I. Shlyakhter, E. B. Hostetter, P. C. Sabeti, C. A. Adebamowo, M. W. Foster, D. R. Gordon, J. Licinio, M. C. Manca, P. A. Marshall, I. Matsuda, D. Ngare, V. O. Wang, D. Reddy, C. N. Rotimi, C. D. Royal, R. R. Sharp, C. Zeng, L. D. Brooks and J. E. McEwen. 2010. “Integrating Common and Rare Genetic Variation in Diverse Human Populations.” *Nature* 467(7311):52–8.

Jakobsson, M., S. W. Scholz, P. Scheet, J. R. Gibbs, J. M. VanLiere, H. C. Fung, Z. A. Szpiech, J. H. Degnan, K. Wang, R. Guerreiro, J. M. Bras, J. C. Schymick, D. G. Hernandez, B. J. Traynor, J. Simon-Sanchez, M. Matarin, A. Britton, J. van de

Leemput, I. Rafferty, M. Bucan, H. M. Cann, J. A. Hardy, N. A. Rosenberg and A. B. Singleton. 2008. “Genotype, Haplotype and Copy-Number Variation in Worldwide Human Populations.” *Nature* 451(7181):998–1003.

Jeong, C., O. Balanovsky, E. Lukianova, N. Kahbatkyzy, P. Flegontov, V. Zaporozhchenko, A. Immel, C. C. Wang, O. Ixan, E. Khussainova, B. Bekmanov, V. Zaibert, M. Lavryashina, E. Pocheshkhova, Y. Yusupov, A. Agdzhoyan, S. Koshel, A. Bukin, P. Nymadawa, S. Turdikulova, D. Dalimova, M. Churnosov, R. Skhalyakho, D. Daragan, Y. Bogunov, A. Bogunova, A. Shtrunov, N. Dubova, M. Zhabagin, L. Yepiskoposyan, V. Churakov, N. Pislegin, L. Damba, L. Saroyants, K. Dibirova, L. Atramentova, O. Utevska, E. Idrisov, E. Kamenshchikova, I. Evseeva, M. Metspalu, A. K. Outram, M. Robbeets, L. Djansugurova, E. Balanovska, S. Schiffels, W. Haak, D. Reich and J. Krause. 2019. “The Genetic History of Admixture across Inner Eurasia.” *Nat Ecol Evol*.

Jobling, MA., E. Hollox, ME. Hurles, T. Kivisild and C. Tyler-Smith. 2014a. *Human Evolutionary Genetics*. New York and Abingdon, United Kingdom: Garland Science.

Jobling, MA., E. Holox, ME. Hurles, T. Kivisild and C. Tyler-Smith. 2014b. “Making Inferences from Diversity.” Pp. 167–223 in *Human Evolutionary Genetics*, edited by G. Science. New York and Abingdon, United Kingdom: Garland Science, Taylor & Francis Group.

Karakachoff, M., N. Duforet-Frebourg, F. Simonet, S. Le Scouarnec, N. Pellen, S. Lecointe, E. Charpentier, F. Gros, S. Cauchi, P. Froguel, N. Copin, D. E. S. I. R. Study Group, T. Le Tourneau, V. Probst, H. Le Marec, S. Molinaro, B. Balkau, R. Redon, J. J. Schott, M. G. Blum, C. Dina and D. E. S. I. R. Study Group. 2015. “Fine-Scale Human Genetic Structure in Western France.” *Eur J Hum Genet* 23(6):831–6.

Karmin, M., L. Saag, M. Vicente, M. A. Wilson Sayres, M. Jarve, U. G. Talas, S. Rootsi, A. M. Ilumae, R. Magi, M. Mitt, L. Pagani, T. Puurand, Z. Faltyskova, F. Clemente, A. Cardona, E. Metspalu, H. Sahakyan, B. Yunusbayev, G. Hudjashov, M. DeGiorgio, E. L. Loogvali, C. Eichstaedt, M. Eelmets, G. Chaubey, K. Tambets, S. Litvinov, M. Mormina, Y. Xue, Q. Ayub, G. Zoraqi, T. S. Korneliussen, F. Akhatova, J. Lachance, S. Tishkoff, K. Momynaliev, F. X. Ricaut, P. Kusuma, H. Razafindrazaka, D. Pierron, M. P. Cox, G. N. Sultana, R. Willerslev, C. Muller, M. Westaway, D. Lambert, V. Skaro, L. Kovacevic, S. Turdikulova, D. Dalimova, R. Khusainova, N. Trofimova, V. Akhmetova, I. Khidiyatova, D. V. Lichman, J. Isakova, E. Pocheshkhova, Z. Sabitov, N. A. Barashkov, P. Nymadawa, E. Mihailov, J. W. Seng, I. Evseeva, A. B. Migliano, S. Abdullah, G. Andriadze, D. Primorac, L. Atramentova, O. Utevska, L. Yepiskoposyan, D. Marjanovic, A. Kushniarevich, D. M. Behar, C. Gilissen, L. Vissers, J. A. Veltman, E. Balanovska, M. Derenko, B. Malyarchuk, A. Metspalu, S. Fedorova, A. Eriksson, A. Manica, F. L. Mendez, T. M. Karafet, K. R. Veeramah, N. Bradman, M. F. Hammer, L. P. Osipova, O. Balanovsky, E. K. Khusnutdinova, K. Johnsen, M. Remm, M. G. Thomas, C. Tyler-Smith, P. A. Underhill, E. Willerslev, R. Nielsen, M. Metspalu, R. Villems and T. Kivisild. 2015. “A Recent Bottleneck of Y Chromosome Diversity Coincides with a Global Change in Culture.” *Genome Res* 25(4):459–66.

Kashimada, K. and P. Koopman. 2010. “Sry: The Master Switch in Mammalian Sex Determination.” *Development* 137(23):3921–30.

Kidd, J. M., N. Sampas, F. Antonacci, T. Graves, R. Fulton, H. S. Hayden, C. Alkan, M. Malig, M. Ventura, G. Giannuzzi, J. Kallicki, P. Anderson, A. Tsalenko, N. A.

Yamada, P. Tsang, R. Kaul, R. K. Wilson, L. Bruhn and E. E. Eichler. 2010. "Characterization of Missing Human Genome Sequences and Copy-Number Polymorphic Insertions." *Nat Methods* 7(5):365–71.

Kimura, M. 1968. "Evolutionary Rate at the Molecular Level." *Nature* 217(129):624–6.

Kingman, J. F. 2000. "Origins of the Coalescent. 1974–1982." *Genetics* 156(4):1461–3.

Kingman, J.F.C. 1982a. "The Coalescent." *Stochastic Process. Appl.* 13:235–48.

Kingman, J.F.C. 1982b. "On the Genealogy of Large Populations." *J. Appl. Probab* 19A:27–43.

Kivisild, T., P. Shen, D. P. Wall, B. Do, R. Sung, K. Davis, G. Passarino, P. A. Underhill, C. Scharfe, A. Torroni, R. Scozzari, D. Modiano, A. Coppa, P. de Knijff, M. Feldman, L. L. Cavalli-Sforza and P. J. Oefner. 2006. "The Role of Selection in the Evolution of Human Mitochondrial Genomes." *Genetics* 172(1):373–87.

Krings, M., A. Stone, R. W. Schmitz, H. Krainitzki, M. Stoneking and S. Paabo. 1997. "Neandertal DNA Sequences and the Origin of Modern Humans." *Cell* 90(1):19–30.

Krings, M., H. Geisert, R. W. Schmitz, H. Krainitzki and S. Paabo. 1999. "DNA Sequence of the Mitochondrial Hypervariable Region Ii from the Neandertal Type Specimen." *Proc Natl Acad Sci U S A* 96(10):5581–5.

Kumar, S. 2005. "Molecular Clocks: Four Decades of Evolution." *Nat Rev Genet* 6(8):654–62.

Lahn, B. T. and D. C. Page. 1999. "Four Evolutionary Strata on the Human X Chromosome." *Science* 286(5441):964–7.

Lawson, D. J., G. Hellenthal, S. Myers and D. Falush. 2012. "Inference of Population Structure Using Dense Haplotype Data." *PLoS Genet* 8(1):e1002453.

Lazaridis, I., N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu, K. I. Bos, S. Nordenfelt, H. Li, C. de Filippo, K. Prufer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fornander, N. Rohland, D. Delsate, M. Francken, J. M. Guinet, J. Wahl, G. Ayodo, H. A. Babiker, G. Bailliet, E. Balanovska, O. Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. M. Bravi, F. Brisighelli, G. B. Busby, F. Cali, M. Churnosov, D. E. Cole, D. Corach, L. Damba, G. van Driem, S. Dryomov, J. M. Dugoujon, S. A. Fedorova, I. Gallego Romero, M. Gubina, M. Hammer, B. M. Henn, T. Hervig, U. Hodoglugil, A. R. Jha, S. Karachanak-Yankova, R. Khusainova, E. Khusnutdinova, R. Kittles, T. Kivisild, W. Klitz, V. Kucinskas, A. Kushniarevich, L. Laredj, S. Litvinov, T. Loukidis, R. W. Mahley, B. Melegh, E. Metspalu, J. Molina, J. Mountain, K. Nakkalajarvi, D. Nesheva, T. Nyambo, L. Osipova, J. Parik, F. Platonov, O. Posukh, V. Romano, F. Rothhammer, I. Rudan, R. Ruizbakiev, H. Sahakyan, A. Sajantila, A. Salas, E. B. Starikovskaya, A. Tarekegn, D. Toncheva, S. Turdikulova, I. Uktveryte, O. Utevska, R. Vasquez, M. Villena, M. Voevoda, C. A. Winkler, L. Yepiskoposyan, P. Zalloua, T. Zemunik, A. Cooper, C. Capelli, M. G. Thomas, A. Ruiz-Linares, S. A. Tishkoff, L. Singh, K. Thangaraj, R. Villems, D. Comas, R. Sukernik, M. Metspalu, M. Meyer, E. E. Eichler, J. Burger, M. Slatkin, S. Paabo, J. Kelso, D. Reich and J. Krause. 2014. "Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans." *Nature* 513(7518):409–13.

Leslie, S., B. Winney, G. Hellenthal, D. Davison, A. Boumertit, T. Day, K. Hutnik, E. C. Royrvik, B. Cunliffe, Consortium Wellcome Trust Case Control, Consortium International Multiple Sclerosis Genetics, D. J. Lawson, D. Falush, C. Freeman, M.

Pirinen, S. Myers, M. Robinson, P. Donnelly and W. Bodmer. 2015. “The Fine-Scale Genetic Structure of the British Population.” *Nature* 519(7543):309–14.

Li, J. Z., D. M. Absher, H. Tang, A. M. Southwick, A. M. Casto, S. Ramachandran, H. M. Cann, G. S. Barsh, M. Feldman, L. L. Cavalli-Sforza and R. M. Myers. 2008. “Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation.” *Science* 319(5866):1100–4.

Li, M., A. Schonberg, M. Schaefer, R. Schroeder, I. Nasidze and M. Stoneking. 2010. “Detecting Heteroplasmy from High-Throughput Sequencing of Complete Human Mitochondrial DNA Genomes.” *Am J Hum Genet* 87(2):237–49.

Libert, F., P. Cochaux, G. Beckman, M. Samson, M. Aksenova, A. Cao, A. Czeizel, M. Claustres, C. de la Rua, M. Ferrari, C. Ferrec, G. Glover, B. Grinde, S. Guran, V. Kucinskas, J. Lavinha, B. Mercier, G. Ogur, L. Peltonen, C. Rosatelli, M. Schwartz, V. Spitsyn, L. Timar, L. Beckman, M. Parmentier and G. Vassart. 1998. “The Deltaccr5 Mutation Conferring Protection against Hiv-1 in Caucasian Populations Has a Single and Recent Origin in Northeastern Europe.” *Hum Mol Genet* 7(3):399–406.

Loogvali, E. L., T. Kivisild, T. Margus and R. Villems. 2009. “Explaining the Imperfection of the Molecular Clock of Hominid Mitochondria.” *PLoS One* 4(12):e8260.

Luo, S., C. A. Valencia, J. Zhang, N. C. Lee, J. Slone, B. Gui, X. Wang, Z. Li, S. Dell, J. Brown, S. M. Chen, Y. H. Chien, W. L. Hwu, P. C. Fan, L. J. Wong, P. S. Atwal and T. Huang. 2018. “Biparental Inheritance of Mitochondrial DNA in Humans.” *Proc Natl Acad Sci U S A* 115(51):13039–44.

Luo, S., C. A. Valencia, J. Zhang, N. C. Lee, J. Slone, B. Gui, X. Wang, Z. Li, S. Dell, J. Brown, S. M. Chen, Y. H. Chien, W. L. Hwu, P. C. Fan, L. J. Wong, P. S. Atwal and T. Huang. 2019. “Reply to Lutz-Bonengel Et Al.: Biparental Mtdna Transmission Is Unlikely to Be the Result of Nuclear Mitochondrial DNA Segments.” *Proc Natl Acad Sci U S A* 116(6):1823–24.

Lutz-Bonengel, S. and W. Parson. 2019. “No Further Evidence for Paternal Leakage of Mitochondrial DNA in Humans Yet.” *Proc Natl Acad Sci U S A* 116(6):1821–22.

Maca-Meyer, N., A. M. Gonzalez, J. Pestano, C. Flores, J. M. Larruga and V. M. Cabrera. 2003. “Mitochondrial DNA Transit between West Asia and North Africa Inferred from U6 Phylogeography.” *BMC Genet* 4(1):15.

Mallick, S., H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G. Poletti, F. Balloux, G. van Driem, P. de Knijff, I. G. Romero, A. R. Jha, D. M. Behar, C. M. Bravi, C. Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M. S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C. Jeong, E. B. Starikovskaya, E. Metspalu, J. Parik, R. Villems, B. M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J. T. Wee, R. Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D. Comas, M. F. Hammer, T. Kivisild, W. Klitz, C. A. Winkler, D. Labuda, M. Bamshad, L. B. Jorde, S. A. Tishkoff, W. S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K. Thangaraj, S. Paabo, J. Kelso, N. Patterson and D. Reich. 2016. “The Simons Genome Diversity Project: 300 Genomes from 142 Diverse Populations.” *Nature* 538(7624):201–06.

Margoliash, E. 1963. “Primary Structure and Evolution of Cytochrome C.” *Proc Natl Acad Sci U S A* 50:672–9.

Martinez-Cruz, B., C. Harmant, D. E. Platt, W. Haak, J. Manry, E. Ramos-Luis, D. F. Soria-Hernanz, F. Bauduer, J. Salaberria, B. Oyharcabal, L. Quintana-Murci, D. Comas and Consortium Genographic. 2012. "Evidence of Pre-Roman Tribal Genetic Structure in Basques from Uniparentally Inherited Markers." *Mol Biol Evol* 29(9):2211–22.

Mason, P. A., E. C. Matheson, A. G. Hall and R. N. Lightowlers. 2003. "Mismatch Repair Activity in Mammalian Mitochondria." *Nucleic Acids Res* 31(3):1052–8.

McVean, G. 2009. "A Genealogical Interpretation of Principal Components Analysis." *PLoS Genet* 5(10):e1000686.

Merzoug, S. and L. Sari. 2008. "Re-Examination of the Zone I Material from Tamar Hat (Algeria): Zooarchaeological and Technofunctional Analyses." *African Archaeological Review* 25(1):57–73.

Mishmar, D., E. Ruiz-Pesini, P. Golik, V. Macaulay, A. G. Clark, S. Hosseini, M. Brandon, K. Easley, E. Chen, M. D. Brown, R. I. Sukernik, A. Olckers and D. C. Wallace. 2003. "Natural Selection Shaped Regional Mtdna Variation in Humans." *Proc Natl Acad Sci USA* 100(1):171–76.

Moilanen, J. S. and K. Majamaa. 2003. "Phylogenetic Network and Physicochemical Properties of Nonsynonymous Mutations in the Protein-Coding Genes of Human Mitochondrial DNA." *Mol Biol Evol* 20(8):1195–210.

Ngamphiw, C., A. Assawamakin, S. Xu, P. J. Shaw, J. O. Yang, H. Ghang, J. Bhak, E. Liu, S. Tongsima and Hugo Pan-Asian SNP Consortium. 2011. "Pansnpdb: The Pan-Asian Snp Genotyping Database." *PLoS One* 6(6):e21451.

Nielsen, R. and M. A. Beaumont. 2009. "Statistical Inferences in Phylogeography." *Mol Ecol* 18(6):1034–47.

Nordborg, M. 2001. "Coalescent Theory." in *Handbook of Statistical Genetics*, edited by D. Balding, Bishop, M., Cannings, C. Chichester, UK: Wiley.

Novembre, J., T. Johnson, K. Bryc, Z. Kutalik, A. R. Boyko, A. Auton, A. Indap, K. S. King, S. Bergmann, M. R. Nelson, M. Stephens and C. D. Bustamante. 2008. "Genes Mirror Geography within Europe." *Nature* 456(7218):98–101.

Novembre, J. and M. Stephens. 2008. "Interpreting Principal Component Analyses of Spatial Population Genetic Variation." *Nat Genet* 40(5):646–9.

Olivieri, A., A. Achilli, M. Pala, V. Battaglia, S. Fornarino, N. Al-Zahery, R. Scozzari, F. Cruciani, D. M. Behar, J. M. Dugoujon, C. Coudray, A. S. Santachiara-Benerecetti, O. Semino, H. J. Bandelt and A. Torroni. 2006. "The Mtdna Legacy of the Levantine Early Upper Palaeolithic in Africa." *Science* 314(5806):1767–70.

Pagani, L., T. Kivisild, A. Tarekegn, R. Ekong, C. Plaster, I. Gallego Romero, Q. Ayub, S. Q. Mehdi, M. G. Thomas, D. Luiselli, E. Bekele, N. Bradman, D. J. Balding and C. Tyler-Smith. 2012. "Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool." *Am J Hum Genet* 91(1):83–96.

Pagani, L., D. J. Lawson, E. Jagoda, A. Morseburg, A. Eriksson, M. Mitt, F. Clemente, G. Hudjashov, M. DeGiorgio, L. Saag, J. D. Wall, A. Cardona, R. Magi, M. A. Sayres, S. Kaewert, C. Inchley, C. L. Scheib, M. Jarve, M. Karmin, G. S. Jacobs, T. Antao, F. M. Iliescu, A. Kushniarevich, Q. Ayub, C. Tyler-Smith, Y. Xue, B. Yunusbayev, K. Tambets, C. B. Mallick, L. Saag, E. Pocheshkhova, G. Andriadze, C. Muller, M. C. Westaway, D. M. Lambert, G. Zoraqi, S. Turdikulova, D. Dalimova, Z. Sabitov, G. N. Sultana, J. Lachance, S. Tishkoff, K. Momynaliev, J. Isakova, L. D. Damba, M. Gubina, P. Nymadawa, I. Evseeva, L. Atramentova, O. Utevska, F. X. Ricaut, N. Brucato, H. Sudoyo, T. Letellier, M. P. Cox, N. A.

Barashkov, V. Skaro, L. Mulahasanovic, D. Primorac, H. Sahakyan, M. Mormina, C. A. Eichstaedt, D. V. Lichman, S. Abdullah, G. Chaubey, J. T. Wee, E. Mihailov, A. Karunas, S. Litvinov, R. Khusainova, N. Ekomasova, V. Akhmetova, I. Khidiyatova, D. Marjanovic, L. Yepiskoposyan, D. M. Behar, E. Balanovska, A. Metspalu, M. Derenko, B. Malyarchuk, M. Voevoda, S. A. Fedorova, L. P. Osipova, M. M. Lahr, P. Gerbault, M. Leavesley, A. B. Migliano, M. Petraglia, O. Balanovsky, E. K. Khusnutdinova, E. Metspalu, M. G. Thomas, A. Manica, R. Nielsen, R. Villems, E. Willerslev, T. Kivisild and M. Metspalu. 2016. "Genomic Analyses Inform on Migration Events During the Peopling of Eurasia." *Nature* 538(7624):238–42.

Palencia-Madrid, L., S. Cardoso, C. Keyser, J. C. Lopez-Quintana, A. Guenaga-Lizasu and M. M. de Pancorbo. 2017. "Ancient Mitochondrial Lineages Support the Prehistoric Maternal Root of Basques in Northern Iberian Peninsula." *Eur J Hum Genet* 25(5):631–36.

Peng, M. S., W. Xu, J. J. Song, X. Chen, X. Sulaiman, L. Cai, H. Q. Liu, S. F. Wu, Y. Gao, N. T. Abdulloevich, M. E. Afanasevna, K. B. Ibrohimovich, X. Chen, W. K. Yang, M. Wu, G. M. Li, X. Y. Yang, A. Rakha, Y. G. Yao, H. Upur and Y. P. Zhang. 2018. "Mitochondrial Genomes Uncover the Maternal History of the Pamir Populations." *Eur J Hum Genet* 26(1):124–36.

Penny, D., M. D. Hendy and M. A. Steel. 1992. "Progress with Methods for Constructing Evolutionary Trees." *Trends Ecol Evol* 7(3):73–9.

Penny, D., M. D. Hendy and B.R. Holland. 2007. "Phylogenetics: Parsimony, Networks, and Distance Methods." Pp. pp. 489–532 in *Handbook of Statistical Genetics*, edited by D. Balding, M. Bishop and C. Cannings. Chichester: John Wiley & Sons.

Pereira, L., N. M. Silva, R. Franco-Duarte, V. Fernandes, J. B. Pereira, M. D. Costa, H. Martins, P. Soares, D. M. Behar, M. B. Richards and V. Macaulay. 2010. "Population Expansion in the North African Late Pleistocene Signalled by Mitochondrial DNA Haplogroup U6." *BMC Evol Biol* 10:390.

Peter, B. M. 2016. "Admixture, Population Structure, and F-Statistics." *Genetics* 202(4):1485–501.

Pinz, K. G. and D. F. Bogenhagen. 1998. "Efficient Repair of Abasic Sites in DNA by Mitochondrial Enzymes." *Mol Cell Biol* 18(3):1257–65.

Poznik, G. D., Y. Xue, F. L. Mendez, T. F. Willems, A. Massaia, M. A. Wilson Sayres, Q. Ayub, S. A. McCarthy, A. Narechania, S. Kashin, Y. Chen, R. Banerjee, J. L. Rodriguez-Flores, M. Cerezo, H. Shao, M. Gymrek, A. Malhotra, S. Louzada, R. Desalle, G. R. Ritchie, E. Cerveira, T. W. Fitzgerald, E. Garrison, A. Marcketta, D. Mittelman, M. Romanovitch, C. Zhang, X. Zheng-Bradley, G. R. Abecasis, S. A. McCarroll, P. Flicek, P. A. Underhill, L. Coin, D. R. Zerbino, F. Yang, C. Lee, L. Clarke, A. Auton, Y. Erlich, R. E. Handsaker, Consortium Genomes Project, C. D. Bustamante and C. Tyler-Smith. 2016. "Punctuated Bursts in Human Male Demography Inferred from 1,244 Worldwide Y-Chromosome Sequences." *Nat Genet* 48(6):593–9.

Pritchard, J. K., M. Stephens and P. Donnelly. 2000. "Inference of Population Structure Using Multilocus Genotype Data." *Genetics* 155(2):945–59.

Pybus, O. G. and A. Rambaut. 2002. "Genie: Estimating Demographic History from Molecular Phylogenies." *Bioinformatics* 18(10):1404–5.

Pyle, A., G. Hudson, I. J. Wilson, J. Coxhead, T. Smertenko, M. Herbert, M. Santibanez-Koref and P. F. Chinnery. 2015. "Extreme-Depth Re-Sequencing of

Mitochondrial DNA Finds No Evidence of Paternal Transmission in Humans." *PLoS Genet* 11(5):e1005040.

Rasmussen, M. D., M. J. Hubisz, I. Gronau and A. Siepel. 2014. "Genome-Wide Inference of Ancestral Recombination Graphs." *PLoS Genet* 10(5):e1004342.

Reich, D., R. E. Green, M. Kircher, J. Krause, N. Patterson, E. Y. Durand, B. Viola, A. W. Briggs, U. Stenzel, P. L. Johnson, T. Maricic, J. M. Good, T. Marques-Bonet, C. Alkan, Q. Fu, S. Mallick, H. Li, M. Meyer, E. E. Eichler, M. Stoneking, M. Richards, S. Talamo, M. V. Shunkov, A. P. Derevianko, J. J. Hublin, J. Kelso, M. Slatkin and S. Paabo. 2010. "Genetic History of an Archaic Hominin Group from Denisova Cave in Siberia." *Nature* 468(7327):1053–60.

Richards, M., H. Corte-Real, P. Forster, V. Macaulay, H. Wilkinson-Herbots, A. Demaine, S. Papiha, R. Hedges, H.-J. Bandelt and B. Sykes. 1996. "Paleolithic and Neolithic Lineages in the European Mitochondrial Gene Pool." *Am J Hum Genet* 59(1):185–203.

Richards, M., V. Macaulay, E. Hickey, E. Vega, B. Sykes, V. Guida, C. Rengo, D. Sellitto, F. Cruciani, T. Kivisild, R. Villems, M. Thomas, S. Rychkov, O. Rychkov, Y. Rychkov, M. Golge, D. Dimitrov, E. Hill, D. Bradley, V. Romano, F. Cali, G. Vona, A. Demaine, S. Papiha, C. Triantaphyllidis and G. Stefanescu. 2000. "Tracing European Founder Lineages in the near Eastern Mtdna Pool." *Am J Hum Genet* 67(5):1251–76.

Rossi, M. N., M. Carbone, C. Mostocotto, C. Mancone, M. Tripodi, R. Maione and P. Amati. 2009. "Mitochondrial Localization of Parp-1 Requires Interaction with Mitofilin and Is Involved in the Maintenance of Mitochondrial DNA Integrity." *J Biol Chem* 284(46):31616–24.

Rowold, D. J., J. R. Luis, M. C. Terreros and R. J. Herrera. 2007. "Mitochondrial DNA Geneflow Indicates Preferred Usage of the Levant Corridor over the Horn of Africa Passageway." *J Hum Genet* 52(5):436–47.

Ruiz-Pesini, E., D. Mishmar, M. Brandon, V. Procaccio and D. C. Wallace. 2004. "Effects of Purifying and Adaptive Selection on Regional Variation in Human Mtdna." *Science* 303(5655):223–6.

Santos, C., R. Montiel, A. Arruda, L. Alvarez, M. P. Aluja and M. Lima. 2008. "Mutation Patterns of Mtdna: Empirical Inferences for the Coding Region." *BMC Evol Biol* 8:167.

Scally, A. and R. Durbin. 2012. "Revising the Human Mutation Rate: Implications for Understanding Human Evolution." *Nat Rev Genet* 13(10):745–53.

Schwartz, M. and J. Vissing. 2002. "Paternal Inheritance of Mitochondrial DNA." *N Engl J Med* 347(8):576–80.

Schwartz, M. and J. Vissing. 2004. "No Evidence for Paternal Inheritance of Mtdna in Patients with Sporadic Mtdna Mutations." *J Neurol Sci* 218(1–2):99–101.

Secher, B., R. Fregel, J. M. Larruga, V. M. Cabrera, P. Endicott, J. J. Pestano and A. M. Gonzalez. 2014. "The History of the North African Mitochondrial DNA Haplogroup U6 Gene Flow into the African, Eurasian and American Continents." *BMC Evol Biol* 14:109.

Sekido, R. and R. Lovell-Badge. 2008. "Sex Determination Involves Synergistic Action of Sry and Sf1 on a Specific Sox9 Enhancer." *Nature* 453(7197):930–4.

Sigurgardottir, S., A. Helgason, J. R. Gulcher, K. Stefansson and P. Donnelly. 2000. "The Mutation Rate in the Human Mtdna Control Region." *Am J Hum Genet* 66(5):1599–609.

Silva, M., M. Oliveira, D. Vieira, A. Brandao, T. Rito, J. B. Pereira, R. M. Fraser, B. Hudson, F. Gandini, C. Edwards, M. Pala, J. Koch, J. F. Wilson, L. Pereira, M. B. Richards and P. Soares. 2017. "A Genetic Chronology for the Indian Subcontinent Points to Heavily Sex-Biased Dispersals." *BMC Evol Biol* 17(1):88.

Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier, L. G. Brown, S. Repping, T. Pyntikova, J. Ali, T. Bieri, A. Chinwalla, A. Delehaunty, K. Delehaunty, H. Du, G. Fewell, L. Fulton, R. Fulton, T. Graves, S. F. Hou, P. Latrielle, S. Leonard, E. Mardis, R. Maupin, J. McPherson, T. Miner, W. Nash, C. Nguyen, P. Ozersky, K. Pepin, S. Rock, T. Rohlfing, K. Scott, B. Schultz, C. Strong, A. Tin-Wollam, S. P. Yang, R. H. Waterston, R. K. Wilson, S. Rozen and D. C. Page. 2003. "The Male-Specific Region of the Human Y Chromosome Is a Mosaic of Discrete Sequence Classes." *Nature* 423(6942):825–37.

Skoglund, P. and I. Mathieson. 2018. "Ancient Genomics of Modern Humans: The First Decade." *Annu Rev Genomics Hum Genet* 19:381–404.

Soares, P., L. Ermini, N. Thomson, M. Mormina, T. Rito, A. Rohl, A. Salas, S. Oppenheimer, V. Macaulay and M. B. Richards. 2009. "Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock." *Am J Hum Genet* 84(6):740–59.

Stringer, C. B. and P. Andrews. 1988. "Genetic and Fossil Evidence for the Origin of Modern Humans." *Science* 239(4845):1263–8.

Sudmant, P. H., T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, K. Ye, G. Jun, M. H. Fritz, M. K. Konkel, A. Malhotra, A. M. Stutz, X. Shi, F. P. Casale, J. Chen, F. Hormozdiari, G. Dayama, K. Chen, M. Malig, M. J. P. Chaisson, K. Walter, S. Meiers, S. Kashin, E. Garrison, A. Auton, H. Y. K. Lam, X. J. Mu, C. Alkan, D. Antaki, T. Bae, E. Cerveira, P. Chines, Z. Chong, L. Clarke, E. Dal, L. Ding, S. Emery, X. Fan, M. Gujral, F. Kahveci, J. M. Kidd, Y. Kong, E. W. Lameijer, S. McCarthy, P. Flicek, R. A. Gibbs, G. Marth, C. E. Mason, A. Menelaou, D. M. Muzny, B. J. Nelson, A. Noor, N. F. Parrish, M. Pendleton, A. Quitadamo, B. Raeder, E. E. Schadt, M. Romanovitch, A. Schlattl, R. Sebra, A. A. Shabalin, A. Untergasser, J. A. Walker, M. Wang, F. Yu, C. Zhang, J. Zhang, X. Zheng-Bradley, W. Zhou, T. Zichner, J. Sebat, M. A. Batzer, S. A. McCarroll, Consortium Genomes Project, R. E. Mills, M. B. Gerstein, A. Bashir, O. Stegle, S. E. Devine, C. Lee, E. E. Eichler and J. O. Korbel. 2015. "An Integrated Map of Structural Variation in 2,504 Human Genomes." *Nature* 526(7571):75–81.

Szczesny, B., A. W. Tann, M. J. Longley, W. C. Copeland and S. Mitra. 2008. "Long Patch Base Excision Repair in Mammalian Mitochondrial Genomes." *J Biol Chem* 283(39):26349–56.

Tajima, F. 1983. "Evolutionary Relationship of DNA Sequences in Finite Populations." *Genetics* 105(2):437–60.

Tamura, K. and M. Nei. 1993. "Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees..". *Mol. Biol. Evol.* 10:512–26.

Tang, H., J. Peng, P. Wang and N. J. Risch. 2005. "Estimation of Individual Admixture: Analytical and Study Design Considerations." *Genet Epidemiol* 28(4):289–301.

Taylor, R. W., M. T. McDonnell, E. L. Blakely, P. F. Chinnery, G. A. Taylor, N. Howell, M. Zeviani, E. Briem, F. Carrara and D. M. Turnbull. 2003. "Genotypes from Patients Indicate No Paternal Mitochondrial DNA Contribution." *Ann Neurol* 54(4):521–4.

Templeton, A. R. 1998. “Nested Clade Analyses of Phylogeographic Data: Testing Hypotheses About Gene Flow and Population History.” *Mol Ecol* 7(4):381–97.

Tilford, C. A., T. Kuroda-Kawaguchi, H. Skaletsky, S. Rozen, L. G. Brown, M. Rosenberg, J. D. McPherson, K. Wylie, M. Sekhon, T. A. Kucaba, R. H. Waterston and D. C. Page. 2001. “A Physical Map of the Human Y Chromosome.” *Nature* 409(6822):943–5.

Torroni, A., T. G. Schurr, M. F. Cabell, M. D. Brown, J. V. Neel, M. Larsen, D. G. Smith, C. M. Vullo and D. C. Wallace. 1993a. “Asian Affinities and Continental Radiation of the Four Founding Native American Mtdnas.” *Am J Hum Genet* 53(3):563–90.

Torroni, A., R. I. Sukernik, T. G. Schurr, Y. B. Starikorskaya, M. F. Cabell, M. H. Crawford, A. G. Comuzzie and D. C. Wallace. 1993b. “Mtdna Variation of Aboriginal Siberians Reveals Distinct Genetic Affinities with Native Americans.” *Am J Hum Genet* 53(3):591–608.

Torroni, A., H. J. Bandelt, V. Macaulay, M. Richards, F. Cruciani, C. Rengo, V. Martinez-Cabrera, R. Villems, T. Kivisild, E. Metspalu, J. Parik, H. V. Tolk, K. Tambets, P. Forster, B. Karger, P. Francalacci, P. Rudan, B. Janicijevic, O. Rickards, M. L. Savontaus, K. Huoponen, V. Laitinen, S. Koivumaki, B. Sykes, E. Hickey, A. Novelletto, P. Moral, D. Sellitto, A. Coppa, N. Al-Zaheri, A. S. Santachiara-Benerecetti, O. Semino and R. Scozzari. 2001. “A Signal, from Human Mtdna, of Postglacial Recolonization in Europe.” *Am J Hum Genet* 69(4):844–52.

Tuppen, H. A., E. L. Blakely, D. M. Turnbull and R. W. Taylor. 2010. “Mitochondrial DNA Mutations and Human Disease.” *Biochim Biophys Acta* 1797(2):113–28.

van de Loosdrecht, M., A. Bouzouggar, L. Humphrey, C. Posth, N. Barton, A. Aximu-Petri, B. Nickel, S. Nagel, E. H. Talbi, M. A. El Hajraoui, S. Amzazi, J. J. Hublin, S. Paabo, S. Schiffels, M. Meyer, W. Haak, C. Jeong and J. Krause. 2018. “Pleistocene North African Genomes Link near Eastern and Sub-Saharan African Human Populations.” *Science* 360(6388):548–52.

van Oven, M. and M. Kayser. 2009. “Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation.” *Hum Mutat* 30(2):E386–94.

Vyas, D. N., A. Kitchen, A. T. Miro-Herrans, L. N. Pearson, A. Al-Meeri and C. J. Mulligan. 2016. “Bayesian Analyses of Yemeni Mitochondrial Genomes Suggest Multiple Migration Events with Africa and Western Eurasia.” *Am J Phys Anthropol* 159(3):382–93.

Wallace, D. C. and D. Chalkia. 2013. “Mitochondrial DNA Genetics and the Heteroplasmy Conundrum in Evolution and Disease.” *Cold Spring Harb Perspect Biol* 5(11):a021220.

Wang, Y. and D. F. Bogenhagen. 2006. “Human Mitochondrial DNA Nucleoids Are Linked to Protein Folding Machinery and Metabolic Enzymes at the Mitochondrial Inner Membrane.” *J Biol Chem* 281(35):25791–802.

Wei, Wei, Salih Tuna, Michael J. Keogh, Katherine R. Smith, Timothy J. Aitman, Phil L. Beales, David L. Bennett, Daniel P. Gale, Maria A. K. Bitner-Glindzicz, Graeme C. Black, Paul Brennan, Perry Elliott, Frances A. Flinter, R. Andres Floto, Henry Houlden, Melita Irving, Ania Koziell, Eamonn R. Maher, Hugh S. Markus, Nicholas W. Morrell, William G. Newman, Irene Roberts, John A. Sayer, Kenneth G. C. Smith, Jenny C. Taylor, Hugh Watkins, Andrew R. Webster, Andrew O. M. Wilkie, Catherine Williamson, Sofie Ashford, Christopher J. Penkett, Kathleen E. Stirrups, Augusto Rendon, Willem H. Ouwehand, John R. Bradley, F. Lucy Raymond, Mark

Caulfield, Ernest Turro and Patrick F. Chinnery. 2019. “Germline Selection Shapes Human Mitochondrial DNA Diversity.” *Science* 364(6442):eaau6520.

Wolpoff, M., X. Z. Wu and A. G. Thorne. 1984. “Modern Homo Sapiens Origins: A General Theory of Hominid Evolutioninvolving the Fossil Evidence from East Asia.” Pp. 411–83 in *The Origin of Modern Humans: A World Survey of the Fossil Evidence*, edited by F. H. Smith and F. Spencer. New York: Alan L. Riss.

Yunusbayev, B., M. Metspalu, E. Metspalu, A. Valeev, S. Litvinov, R. Valiev, V. Akhmetova, E. Balanovska, O. Balanovsky, S. Turdikulova, D. Dalimova, P. Nymadawa, A. Bahmanimehr, H. Sahakyan, K. Tambets, S. Fedorova, N. Barashkov, I. Khidiyatova, E. Mihailov, R. Khusainova, L. Damba, M. Derenko, B. Malyarchuk, L. Osipova, M. Voevoda, L. Yepiskoposyan, T. Kivisild, E. Khusnutdinova and R. Villems. 2015. “The Genetic Legacy of the Expansion of Turkic-Speaking Nomads across Eurasia.” *PLoS Genet* 11(4):e1005068.

Zerjal, T., Y. Xue, G. Bertorelle, R. S. Wells, W. Bao, S. Zhu, R. Qamar, Q. Ayub, A. Mohyuddin, S. Fu, P. Li, N. Yuldasheva, R. Ruzibakiev, J. Xu, Q. Shu, R. Du, H. Yang, M. E. Hurles, E. Robinson, T. Gerelsaikhan, B. Dashnyam, S. Q. Mehdi and C. Tyler-Smith. 2003. “The Genetic Legacy of the Mongols.” *Am J Hum Genet* 72(3):717–21.

Zhivotovsky, L. A., P. A. Underhill, C. Cinnioglu, M. Kayser, B. Morar, T. Kivisild, R. Scozzari, F. Cruciani, G. Destro-Bisol, G. Spedini, G. K. Chambers, R. J. Herrera, K. K. Yong, D. Gresham, I. Tournev, M. W. Feldman and L. Kalaydjieva. 2004. “The Effective Mutation Rate at Y Chromosome Short Tandem Repeats, with Application to Human Population-Divergence Time.” *Am J Hum Genet* 74(1):50–61.

Zuckerkandl, E. and L. Pauling. 1962. “Molecular Disease, Evolution, and Genetic Heterogeneity.” Pp. 189–225 in *Horizons in Biochemistry*, edited by M. Kasha and B. Pullman. New York: Academic Press.

Zuckerkandl, E. and L. Pauling. 1965. “Evolutionary Divergence and Convergence in Proteins.” Pp. pp. 97–166 in *Evolving Genes and Proteins*, edited by V. Bryson and H. J. Vogel. New York: Academic Press.

ACKNOWLEDGEMENTS

Ah, the realm of the theme and variations of “thank you”, a heaven of seemingly less contrived nature, a place probably less skimmed through than the rest of the corpus of a thesis, where expectations of reading your name may, or not…, run high. Given the date of the first reference and the fact that the whole adventure started few years earlier than that, that makes a lot of people to thank for, and a lot of people to forget to do so. I sincerely apologise if I fail to give you below the due credit you feel you deserve. Mark it down simply to forgetfulness, not out of spite or for other ultimate motive of unpleasant nature. Were you to have any complain, please find my details in the CV section and send me a strongly worded letter, UN style, and I will see to fix things (would sending you a very politely worded letter be suitable?). Anyway, let’s dive in.

In a completely expected fashion, I shall start with my seminal supervisor, Professor Richard Villems, whose patience is rivalled only by his knowledge. It has been a privilege to do my research work under your tutelage and I am extremely grateful for the chance you gave me to join your team these, few, years back. I never cease to be amazed by how easy it is to go and ask you questions, which, from a French perspective, at equal position, would usually require making an appointment and sit in the corridor for a while before getting in. Whilst in your office, comfortably ensconced in a chair, it is unfailingly pleasant to be taken on a ride, passing through various places and eras, ending up with a much broader perspective than I had at the beginning.

In a less expected fashion, it is not possible for me not to thank the late Professor Jean-Paul Moisan, whose scientific thirst is the very reason I came to Estonia. Were it not for you, this whole chapter, nay, volume of my life, would not have even existed. I have, so far, always had the chance of having enormously talented supervisors (and colleagues too), and you were the very first. Beaj vat Komonor !

Speaking about lasting impression, how befitting to turn to the quietly spectacular Professor Toomas Kivisild. I wonder what would have happened had the mermaids of the perfidious Albion not succeeded in luring you away from here. Or was it, after all, merely just the normal course of your career? Regardless, the sharpness, cold accuracy; wittiness too, of your opinions and criticisms are something to behold. Taking my modest part in the writing of the Ethiopian paper and benefitting from your guidance for writing my “own” papers was equally a showcase of your scientific prowess. I shall never forget the most splendid grin you were sporting when you broke the news to me that Anna’s paper will be shortly accepted in Science; that was the only magnificent bit of the whole story. Not too sure if I would not exchange experiencing these radiant smile and face for a path to U6 and M1 less cluttered with traps and people having a head-start on me…

Looking back, it was the normal course, the only logical way things could go. From an ever so joyful PhD student ready to partake in and crush the

competition in games of 501 to an ever still so joyful Director of the Institute of Genomics. And through it all, you have always kept this accessibility and readiness to share your knowledge and explain some of the concepts that where challenging my comprehension. Thanks to you and your technical help, my later papers could rely on nicer artworks than before; a touch that surely can help assuaging potential grumpy reviewers, and was useful in making a thesis that much more eye pleasing as well, however thrifty it is in artwork (lack of time maybe…). It is a real shame we never got to publish the expectation and absolute scales we ingeniously devised in Split, humanity would have greatly benefited from them, most assuredly. On a slightly more serious note, I am truly thankful to you for providing the good words and the sterner ones to help seeing this whole project through. And your suggestion of the after-publication service was nothing short of brilliant. And before I forget, thank you very much too for saving me from potential embarrassment by bringing to my attention the various faulty bits in the parts above. And now; listen very carefully, I shall say zis only once: il faut noter les beaux efforts de communication francophone ; il faut persévérer. Ühesõnaga, Merci beaucoup Monsieur Mait.

That concludes the "mandatory" part, although calling it so does not preclude in the least its utter sincerity, nor does it call into question the point of writing what follows. I will now get into the territory where the forgetfulness mentioned above will inevitably start to creep in. It probably will have a heavy chronological flavour, but that is one way to avoid forgetting too many people.

If Jean-Paul and Richard were the ones responsible for me even making it here, the Croatian girls are without a doubt the ones responsible for me even staying here at all… Not because the lab might have been a scary place, but back in late September 2001, for a young French guy who had never exactly traveled abroad, Estonia was surely the "exotic place" Jean-Paul had proposed me to go study in. But Lovorka, Marijana and Svetlana, you made sure to smooth the abrupt transition, sharing your experience of foreign students in Tartu. Without your help, this story would have likely been cut short. In that regard, I also got the chance to play that role a bit when you came here Alexandra, although it might as likely be that the help was more mutual than anything. You have been a blast to have as neighbour, not least because you did my laundry, which was super nice of you.

Logically, if speaking about Alexandra, I inevitably think about you too my dearest Carolina, and not only because your name is a homonym to the dearest Karoliina of all (more on that later). We have had some most excellent time together and your perpetual joie de vivre and relax attitude was contagious. Although it would appear that I am about to defend my PhD, I free you from your promise to get me a pair of John Lobb; but if you were to insist, let me know, my shoe size is still 9 in UK…

Following this thread, first stop outside of the lab biotope, Portuguese girls equal my dear friendo Dirk. Your part in my acclimation to the Estonian environment is unsurpassed. Listening to your recent three months in South-East Asia or your mission in Mexico brought some perspective to what was a

Tartu in the process of blossoming into the much more Western European city-like place it is now. Thanks to you, my musical horizon got to greatly expand in such few short months, discovering gems unknown to my ears. That made looking at electropherograms that much more appealing and pleasant. More importantly, I cannot thank you enough for all the help with learning English and many of its idiomatic expressions, providing an expert hand in fighting my nescience. And because I am not one to shy away in the face of a fat preterition example, it goes without saying that I am very much indebted to your unwavering willingness to provide a shoulder for sharing the burden of some of my existential issues (the pleonasm level of "shouldering" was too much…). I hope I will get a chance to show off my perfect snatches in your home gym... :D

Second batch of Croatian natives. Nina, Jelena, Dubravka and Tena, last shipment of Croatian girls from the dearest Professor Pavao Rudan, but what a selection. What an energetic bunch! Your departure left a resounding void (but more PCR machines available), but that made seeing you again on your home turf even more enjoyable.

Keeping with the Slavic flavour, but closer to "home", the lab has always had a very close relation with Russia. I will inevitably strike the first misses here (although some of the Balkans representatives are already missing, Lejla and co.), and I would like to thank you guys, Ildus, Sergey, Sergey (how likely was it to have at least two Sergeys?), Andrey (just one?), Sardana *et al*. for giving me a starter in the beautiful Russian language. Oleg, it was a real pleasure to meet you again back in May when many of us were treated to a very good conference in wonderful Moscow. It was a great scientific and "extra-scientific" experience. Still in Russia, administratively speaking, I am most grateful to you Bayazit for sharing your expertise in coalescent theory and other related probability subjects, which inevitably led to expanding the horizon of my understanding. Thank you also for having shown me the unexpected ways to apply our usual package of methods to explore questions quite far from our daily inquiries.

The lab got to surprise Toomas for his jubilee this August, and we acknowledged his pivotal role in starting the Indian connection here. I still remember your striking first steps in Estonia Gyaneshwer. Passing from + 28°C to – 26°C during the journey made certain to make it memorable, all wrapped up in a light parka with a measly scarf around your head for lack of a suitable hat. But if your arrival was memorable, it pales in comparison to what it led to. Your friendship is warmer than the hottest day in India and you and Chandana are sweeter than any kulfi. You both are the very perfect embodiment of benevolence. You have always been there, through good and bad times, ready to share the laughs and the burden as the case may be. I truly miss our culinary exchanges, and I am beholden to you for making me realise that a dish does not necessarily need to contain some meat protein; dhal and chana masala for the win! I keep my fingers crossed to finally get to discover your home country someday.

And if Gyaneshwer and Chandana are here, chances are good that you are not too far either Lena. You have one of these amazing personalities, always smiling and happy, you are a person whose sheer presence is enough to give you a boost and you are not stingy in sharing good thoughts!

Good thoughts are a necessity that you possess in spades my dearest Kristiina, and, were you so kind and inclined as to allow me a bit of French banter, you are not short on good looks either… I am truly indebted to your unwavering support during the ups and downs of my scientific endeavours. At times, it might even have been pretty reminiscent of what a mother would have done, which, I am most certainly convinced, mine would have done were it not for her early departure to go and meet the "Grand Nénuphar". I am as well very much indebted to you (and Meelis) for your most wonderful translation work. Without it, this thesis would have stalled short of the finish line. You have also provided plenty of sound scientific advice along the whole way; good that you are my desk-mate, I do not have to go far to get these precious nuggets of knowledge. I am very thankful to you as well for having shared with me your spots of the Estonian nature worthy of a visit in family; fresh air is important for a clear mind. Suur, suur aitäh!

French connection, same place in France, minus the heroin. Professor Chiaroni; mon très cher Jacques, I am very much obliged to you for the great opportunity you gave me to work on the Afghan project. Gracie assai. Besides the stimulating scientific side of the collaboration, that gave me the chance to meet Julie and Stéphane. Thank you, Julie, for your invaluable help, as well as the tasty raviolis. Stéphane, thank you very much for imparting your statistical knowledge, but especially for dealing with that whole bunch in the project… That would be great if we were to strengthen our collaboration, I am convinced we could do some excellent research together; Erika's recent fruitful endeavour (keep going Erika!) is a good example.

Still in France, back in time, on the banks of the Loire, the CHU of Nantes and INSERM U463. The Laboratoire d'Étude du Polymorphisme de l'ADN is being created. I still remember fondly the unceasing barrage of jokes that you were firing at me my dear Sébastien and Stéphane. No ration stamp, no Trabant (not endemic that far east), but some moose for sure! And dropped in the middle of it, you, Chrystelle. We have been together in getting to learn this field new to us. Without your help and immense work, the French study would not have seen the light of day.

To get back on the Afghan subject, it was really on honour to finally work with you Professor Underhill. Although I got to meet you earlier in the wonderful setting of Hvar, it was with the Hindu Kush project that I got to experience first-hand your endless knowledge of that little thing that makes us males. I am very thankful for the words of wisdom and the kind ones you had with regard to me.

Carrying on with the theme of heavy weights of population genetics, it is always a pleasure to meet you Doron. Not least because when it is in Tartu, you inevitably have in tow some gastronomic delicacies; the thought of which

makes me salivate right now. Your readiness to discuss things and get them done is truly impressive. I am ever so thankful to have had the chance to contribute to your staggering number of publications. Without your contribution (and the kind participants who were willing to take part in the scientific project), the paper would have had a narrower scope. [To the editor: Insert "Thank you very much my dear friend" in Hebrew, will you please? ;)]

My dearest chap! I am sincerely embarrassed that you got to experience some of the multifarious unpleasant sides of doing Science in France (never got to myself…), and the other non-scientific bits too (those I most surely did, both at the receiving and at the delivering end…). Or maybe not, because then you were able to bring some dairy awesomeness to Tartu. Pigouille, my tasty little friend… Thank you for your input in improving the more relevant chapters, it seems that in a few cases I could not see the forest for the trees. I sincerely hope that the Breton impromptu side-project will see a fecund ending.

Monika, my dear, dear fellow bookworm, or bibliophile if inclined to go for a less figurative flavour. May we be able to build a proper anti-library, populated by numerous *tsundoku*… The path we followed had quite few similarities, but you got to the deck of the galley (the rowing type, not the place where the tasty treats are kept) earlier. I shall now stop my rowing in its depths very soon, to join you on deck, and I will try my very best to abide by your written "injunction", but that might take a little effort, or not. Regardless, you have been instrumental in speeding up my rowing, and your input on the Y chromosome part helped overcoming my shortcomings in what is a part of me, funnily enough. It was always salutary to discuss with you the hurdles we were facing, since, more often than not, you had already cleared them! I hope your relationship with New-Zealand will be as successful and blessed as it has been for our trail blazer.

Товарищ Георгий! As with every bright mind, it is necessary to get through few peculiar ways before getting to fully appreciate you. You may not be a paragon of equanimity, but that makes getting your point of view that much clearer. You are one of the few people that rather do things in their own way, because you question them and find a way to improve them. Your autonomy and your ability to learn complex things by yourself are nothing short of impressive. You are also without a doubt the only walking and breath holding reason of (a) population decline(s) I have ever met.

Professor Mark G. Thomas, attending your presentation has been an eye-opener, very much akin to a revelation. I realised how much of a sheeple I had become, letting myself follow leisurely what I was learning and doing, having checked in my critical thinking at the entrance. Open the flood gates of scepticism, back to have a more critical outlook of things, not to gobble them up without second thoughts. That requires an open mind, humility, a wealth of knowledge and a healthy pinch of epistemology. That is why I am hurrying to catch up in some fields particularly relevant to becoming better equipped with a healthy and informed sceptical stance, devoid of figurative Manicheism. "In all affairs it's a healthy thing now and then to hang a question mark on the things

you have long taken for granted." Bertrand Arthur William Russell, 3rd Earl Russell, OM FRS.

I would also like to express my sincere gratitude to Antti Sajantila and Tatjana Bergelt. The calm attitude of the professor balanced by the passionate nature of the artist. The conversation we had until the wee hours was extremely powerful, not quite sure if I have yet the mental fortitude and guts to follow through with the realisation I came to…

Much credit and palju "thank you" are due to other wonderful people of the lab. I am very grateful to the "three ladies", Ene, Maere and Siiri. Thank you for your kind help and advice when I started to take my first steps in the lab. Ille, you have been there as well to help me get my bearings in the lab, finding the necessary solution or recipient for making magic happen. Jüri, your biochemical acumen has always been a reliable source of help. I have you to thank for infecting me with the camera bug, but, being in close contact with *el paparazzo*, it was bound to happen at some point. I am also appreciatory of all the technical help and service provided by the competent people of what has become the EBC core lab, especially you Tuuli. Your kind understanding regarding my "urgent" sequences was only partially rewarded by having you on the M1 and U6 journey, but rest assure that your diligence has not gone unnoticed in my books. Anu, thank you for keeping me on my toes and relentlessly testing my Estonian and also for bringing fashion into the little microcosm that is our office (note to the other graceful ladies populating said office; you are not at all badly dressed either). Thank you, Mari, for the sugary fuel for the brain, your bees know their trade, and so do your apple trees. I am amazed by your ability to tackle all your duties and hobbies in what would probably require a few separate lives for most people. Anne-Mai and Ajai, thank you very much for unconsciously breathing down my neck, hastening my pace to wrap things up. I wish you all the best with what is left to be done. As for you, Lehti, I would assume not even the slides are left to be done as of the time of writing these words. Drive and dedication are your middle names, aren't they? I hope you will get to read my most sincere and frank "Merde" on time for your defence. Speaking of dedication, yours Hovannes is also very impressive. I am glad to have had the chance to experience it with your successful U7 project. I am convinced that your J1 J2 [naïve p(being correct) = 0.5; won't take my chances, cross the wrong one ;)] master project will also get a very deserved happy conclusion. Alexander, Alexander, Alexander; thank you for the fleeting moments of blissful escape… It would appear that good nature is some sort of Belarussian trait; who would have guessed? Indeed, Vasili, you are as equally always smiling and happy than Lena is, and that can be contagious. Speaking of smile, yours, Kai, is the prototypical awesome one, so full of life.

I would like also to thank you Professor Metspalu for agreeing to be my internal opponent and constantly inquiring about the advancement of the manuscript. Unfortunately, the advent of the Institute of Genomics disqualified you for this difficult task. That is why I am that much more thankful to you

Professor Remm for stepping in. I will not frame your e-mail, but it is so inside me.

Albeit you guys came a bit later in the game, you nonetheless provided an excellent atmosphere befitting to productive exchanges. Rodrigo, wherever you will be, you will be missed, except if you are still in lab of course… But IT has many good sides, one of them being the ease of communication despite physical distances. If I ever get to explore the Cenotes, I will be sure to check and get myself some *al pastor* goodness. Bringing an unmistakable sunny touch, the lively Italian crew, Francesco (thanks for the continuing education), Davide, Linda and Ludovica, for which we can thank you, Luca. Wikipedia will have to consider in the near future to update the People section of its page about Lucca. The way you approach problems and see things is not without reminding me of a certain Toomas… I am eagerly awaiting to know the fate of the Chargaffity can of worms! Mayukh, I am looking forward to picking your brain further about selection. Burak, I am envious, let's be honest, of the easiness you have in dealing with machine learning. There is much to be learnt but the potential reward in thinking outside of the box with a ginormous box is immense, full of promises. Katri, thank you for taking part in deciphering the $\binom{n}{k}$ mystery. Dankeschön! to our Germanic speakers, Meriam (merci beaucoup marche aussi, non ?), Marcel and Tina, inquiring people of ancient DNA. That would happen to include you too Freddi. The deadpan humour is strong with this one. ☺

Special thanks to the administrative people, without whom the whole edifice of Science would not get very far. Thank you Merilin, Mariza, Tiina (congratulations!) and colleagues at UT, Mrs Doums and Mrs Cardaillac-André at the EPHE for the successful outcome of the battle with the arcana of French bureaucracy. Ольга, спасибо огромное, что присмотрели за моими растениями. Мое рабочее место выглядит замечательно.

Redis Mäger, maybe this will help:

```
perl -p -i -e "s/`erwin/`erwan/g" general_knowledge.brain
```

Thank you Matthieu for showing me the proper way of being thèsard, will try it for the next thesis.

If you have not read your name yet (Lauri, Eva-Liis, Helle-Viivi, Helen, Ivar, Reidar…), sorry, you will need to follow the instructions written above. I could have asked for the list of all current and past "employees", but that would have stretch things a bit…

Pierre-Yves, no problem for you with English. You probably followed this little circus more from the side, but our birthday's conversations were clear enough about your support. I am looking forward to having you around, after all those years of only electronic meetings.

Gwen, même si je suis certain que tu continues à faire tes exercices gymglish avec application (comment pourrait-il en être autrement ?), je te fais grâce de la lecture anglaise pour cette partie à ton attention. Même si nous avons des relations familiales typiquement Pennarun, je ne doute pas un instant que ton soutien fut constant durant toutes ces années. J'en profite aussi pour t'assurer du mien dans ta quête académique ; elles sont bien ardues parfois, surtout quand

certains s'acharnent à se foutre des bâtons dans les roues… La vie, elle, s'est allègrement chargée de t'en filer quelques stères bien pourris. Je ne peux qu'admirer ta force et ta résilience et je souhaite de tout cœur qu'après cette scoumoune ligneuse, elle te réserve des surprises fort plus agréables ! Je suis persuadé que France, Corentin, Johanika et Patxi se chargent de t'en fournir quelques-unes, entre quelques conneries séantes aux enfants bien nés ; et puis Philippe n'est sûrement pas avare à l'occasion de bonnes choses non plus. Si la distance et la rigidité administrative, combinées à ton sens développé et aigu de probité enseignante, font que tu ne seras probablement pas là pour la soutenance, je suis persuadé que je pourrais compter sur tes pensées les plus affectives et ton soutien inconditionnel.

Maman, c'est encore une plaie vive que de n'avoir pu davantage profiter de ta renaissance, tout comme de n'avoir pas pu obtenir mon doctorat avant ton départ pour aller rencontrer le Grand Nénuphar. Je sais que, avec ou sans doctorat, ton amour aurait été le même, mais j'aurais été bien plus fier d'avoir pu au moins réussir cela avant… Même si le temps fait son office, son lent travail de sape, la douleur reste encore bien présente. Mais celle-ci s'efface rapidement lorsque je te revois dessiner ce gros cœur, le dernier, sur la vitre du TGV, parce que je sais que tu seras toujours là pour moi, où que tu sois…

Je ne doute nullement de tes capacités à entraver l'engliche, mais, si Monsieur Père le veut bien, je vais plutôt continuer à tartiner en français. De plus, avec le ton qui sera sûrement un tant soit peu cathartique, Lacan ne serait pas content s'il en était autrement. On peut dire qu'essayer, consciemment ou non, d'émuler le père n'est pas facile et peut amener parfois à des péripéties. Ayant soutenu une thèse en conneries en tout genre depuis fort longtemps déjà, et les études de médecine n'étant pas du tout adaptées à mon style d'apprentissage, il ne restait que ma Science bien aimée pour devenir docteur et prétendre être sur un pied d'égalité. Et me voilà finalement aux portes de ce que je pensais être le Saint Graal. Seulement, je crains que le chemin y conduisant n'ait fortement émoussé l'aura et le prestige que je voulais bien prêter à ce titre. Je ne vais pas renoncer à goûter au plaisir de l'avoir, mais la potion m'a l'air d'être bien plus amère que je ne l'avais imaginée. Que n'eussé-je réalisé plus tôt l'ineptie de croire que seul un tel statut me vaudrait reconnaissance à tes yeux ? La lie, me voilà ! Nous rangerons tout cela dans son tiroir attitré, je finirai le repas totémique et je passerai à l'étape suivante, sachant que ton affection et ta reconnaissance étaient toujours là, j'avais tout simplement oublié de mettre mes lunettes…

In brief, I am forever grateful to you both, Maman, Papa, for always having been there for me and supporting me whatever was the cause of my growing pain. I am really looking up to you, and I thank you for the intellectual atmosphere you created at home, a very stimulating one. Je vous aime.

I have the purest luck of having two most wonderful kids. Thank you Siméon and Madleen for brightening my life. The joy you bring is of quintessential form, so, please, keep on delivering the goods!

If there is one person who has seen it all, it can only be you, Karo. I can hardly express how grateful I am that you weathered all these oft horrible mood swings and their unpleasant externalisation. That you haven't stop believing in me and supporting me is a sheer mystery to me. But, on the other hand, if I were to have ever questioned whether had I been mistaken in being so besotted by you all these years ago, in the light of it all, the clear answer would be a resounding no. Thank you from the bottom of my heart for your loving care and affection. Without them, things would have been inevitably much grimmer. Let us now turn that page and start a whole new chapter, let us now rejoice and relish in the beauty of life, let us celebrate and espouse the glorious moments it is filled with, wherever they may happen to be, however fleeting or lasting they may be, because they are ours.

PUBLICATIONS

CURRICULUM VITAE

Names: Erwan, Pierre, François, Xavier, Marie
Surname: Pennarun
Date of birth: December 15, 1977
Phone: +372 737 5054
E-mail: erwan@ebc.ee

Academic background:

2003–2009	University of Tartu, Estonian Biocentre, PhD student
2001–2003	University of Tartu, Estonian Biocentre, research student
1999–2003	INSERM U463, Nantes, France, EPHE Diploma (MsC) *cum laude*, "Mutational analysis of the 13q14.3 chromosomic region, implicated in the B Chronic Lymphocytic Leukaemia (B-CLL).
1997–1999	I.U.T (Technological University Institute) of Quimper, France, D.U.T. (Technical University Degree) of biological engineering, option Biological and Biochemical Analyses
1996–1997	University of Brest, France, first year of DEUG SVT (Earth and Life Sciences general university studies diploma)

Professional employment:

2018–…	Research fellow at the Estonian Biocentre, Institute of Genomics of the University of Tartu, Estonia.
2009–2017	Research fellow at the Estonian Biocentre at the Evolutionary Biology Department of the Institute of Molecular and Cellular Biology of the University of Tartu, Estonia.
2003–2009	Part-time research fellow at the Estonian Biocentre at the Evolutionary Biology Department of the Institute of Molecular and Cellular Biology of the University of Tartu, Estonia.
2001–2003	Visiting student at the Estonian Biocentre at the Evolutionary Biology Department of the Institute of Molecular and Cellular Biology of the University of Tartu, Estonia.
1999–2001	Technician in INSERM unit U463 in Nantes, France.
Summer 1999	Technician at the Laboratory of Biologic and Biochemical Analyses of the Laennec Hospital in Quimper, France.

Attended courses:

April 2–4, 2019	CodeRefinery workshop, teaching researchers in sustainable software development, Tartu, Estonia
August 27–28, 2019	ELIXIR course on Python, Tartu, Estonia
May 21–22, 2015	ELIXIR R basics course, Tartu, Estonia
June 6–7, 13–14, 2013	Data analysis with Python, Tartu, Estonia
August 22–26, 2013	OpenGENE project's workshop "GWAS: From Genotyping to Sequencing", Tartu, Estonia

May 13–16, 2007 2nd Paris Workshop on Molecular and Statistical Genomic Epidemiology, Paris, France

Publications

Chaubey G, Tamang R, **Pennarun E**, Dubey P, Rai N, Upadhyay RK, Meena RP, Patel JR, van Driem G, Thangaraj K, Metspalu M, Villems R. Reconstructing the population history of the largest tribe of India: the Dravidian speaking Gond. Eur J Hum Genet. 2017 Apr;25(4):493–498. doi: 10.1038/ejhg.2016.198. Epub 2017 Feb 1. Erratum in: Eur J Hum Genet. 2017 Nov;25(11):1291

Sahakyan H, Hooshiar Kashani B, Tamang R, Kushniarevich A, Francis A, Costa MD, Pathak AK, Khachatryan Z, Sharma I, van Oven 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 HV, 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. Origin and spread of human mitochondrial DNA haplogroup U7. Sci Rep. 2017 Apr 7;7:46044.

Di Cristofaro J, **Pennarun E**, Mazières S, Myres NM, Lin AA, Temori SA, Metspalu M, Metspalu E, Witzel M, King RJ, Underhill PA, Villems R, Chiaroni J. Afghan Hindu Kush: where Eurasian sub-continent gene flows converge. PLoS One. 2013 Oct 18;8(10):e76748. doi: 10.1371/journal.pone.0076748. eCollection 2013.

Pennarun E, Kivisild T, Metspalu E, Metspalu M, Reisberg T, Moisan JP, Behar DM, Jones SC, Villems R. Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa. BMC Evol Biol. 2012 Dec 3;12:234. doi: 10.1186/1471-2148-12-234.

Pierron D, Chang I, Arachiche A, Heiske M, Thomas O, Borlin M, **Pennarun E**, Murail P, Thoraval D, Rocher C, Letellier T. Mutation rate switch inside Eurasian mitochondrial haplogroups: impact of selection and consequences for dating settlement in Europe. PLoS One. 2011;6(6):e21543. doi: 10.1371/journal.pone.0021543. Epub 2011 Jun 28.

Ottoni C, Martínez-Labarga C, Loogväli EL, **Pennarun E**, Achilli A, De Angelis F, Trucchi E, Contini I, Biondi G, Rickards O. First genetic insight into Libyan Tuaregs: a maternal perspective. Ann Hum Genet. 2009 Jul;73(Pt 4):438–48. doi:10.1111/j.1469-1809.2009.00526.x. Epub 2009 May 20.

Pierron D, Rocher C, Amati-Bonneau P, Reynier P, Martin-Négrier ML, Allouche S, Batandier C, Mousson de Camaret B, Godinot C, Rotig A, Feldmann D, Bellanne-Chantelot C, Arveiler B, **Pennarun E**, Rossignol R, Crouzet M, Murail P, Thoraval D, Letellier T. New evidence of a mitochondrial genetic background paradox: impact of the J haplogroup on the A3243G mutation. BMC Med Genet. 2008 May 7;9:41. doi: 10.1186/ 1471-2350-9-41.

Richard C, **Pennarun E**, Kivisild T, Tambets K, Tolk HV, Metspalu E, Reidla M, Chevalier S, Giraudet S, Lauc LB, Pericić M, Rudan P, Claustres M, Journel H, Dorval I, Müller C, Villems R, Chaventré A, Moisan JP. An mtDNA perspective of French genetic variation. Ann Hum Biol. 2007 Jan-Feb;34(1):68–79.

Achilli A, Olivieri A, Pala M, Metspalu E, Fornarino S, Battaglia V, Accetturo M, Kutuev I, Khusnutdinova E, **Pennarun E**, Cerutti N, Di Gaetano C, Crobu F, Palli D, Matullo G, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Semino O, Villems R, Bandelt HJ, Piazza A, Torroni A. Mitochondrial DNA variation of modern Tuscans supports the near eastern origin of Etruscans. Am J Hum Genet. 2007 Apr;80(4):759–68. Epub 2007 Feb 6.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, **Pennarun E**, Parik J, Geberhiwot T, Usanga E, Villems R. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet. 2004 Nov;75(5):752–70. Epub 2004 Sep 27. Erratum in: Am J Hum Genet. 2006 Jun;78(6):1097.

Loogväli EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk HV, Parik J, **Pennarun E**, Laos S, Lunkina A, Golubenko M, Barac L, Pericic M, Balanovsky OP, Gusar V, Khusnutdinova EK, Stepanov V, Puzyrev V, Rudan P, Balanovska EV, Grechanina E, Richard C, Moisan JP, Chaventré A, Anagnou NP, Pappa KI, Michalodimitrakis EN, Claustres M, Gölge M, Mikerezi I, Usanga E, Villems R. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. Mol Biol Evol. 2004 Nov;21(11):2012–21. Epub 2004 Jul 14.

Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk HV, Parik J, Loogväli EL, Derenko M, Malyarchuk B, Bermisheva M, Zhadanov S, **Pennarun E**, Gubina M, Golubenko M, Damba L, Fedorova S, Gusar V, Grechanina E, Mikerezi I, Moisan JP, Chaventré A, Khusnutdinova E, Osipova L, Stepanov V, Voevoda M, Achilli A, Rengo C, Rickards O, De Stefano GF, Papiha S, Beckman L, Janicijevic B, Rudan P, Anagnou N, Michalodimitrakis E, Koziel S, Usanga E, Geberhiwot T, Herrnstadt C, Howell N, Torroni A, Villems R. Origin and diffusion of mtDNA haplogroup X. Am J Hum Genet. 2003 Nov;73(5):1178–90. Epub 2003 Oct 20.

Rondeau G, Moreau I, Bézieau S, Petit JL, Heilig R, Fernandez S, **Pennarun E**, Myers JS, Batzer MA, Moisan JP, Devilder MC. Comprehensive analysis of a large genomic sequence at the putative B-cell chronic lymphocytic leukaemia (B-CLL) tumour suppresser gene locus. Mutat Res. 2001 Dec; 458(3–4):55–70.

Bezieau S, Devilder MC, Avet-Loiseau H, Mellerin MP, Puthier D, **Pennarun E**, Rapp MJ, Harousseau JL, Moisan JP, Bataille R. High incidence of N and K-Ras activating mutations in multiple myeloma and primary plasma cell leukemia at diagnosis. Hum Mutat. 2001 Sep;18(3):212–24.

ELULOOKIRJELDUS

Nimi: Erwan, Pierre, François, Xavier, Marie Pennarun
Sunniaeg: 15. detsember 1977
Telefon: +372 737 5054
E-post: erwan@ebc.ee

Haridus:

2003–2009	Tartu Ülikool, Eesti Biokeskus, Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituut, doktorant
2001–2003	Tartu Ülikool, Eesti Biokeskus, külalisüliõpilane
1999–2003	INSERM U463, Nantes, Prantsusmaa, EPHE Diplom (MsC) *cum laude*, "B-kroonilises lümfotsütaarses leukeemias osaleva kromosoomi 13q14.3 mutatsioonianalüüs".
1997–1999	Quimper'i I.U.T (Tehnikaülikooli instituut), Prantsusmaa, D.U.T. (Tehnikaülikooli kraad) biotehnoloogia, Bioloogilised ja biokeemilised analüüsid eriala
1996–1997	Brest'i Ülikool, Prantsusmaa, DEUG SVT esimene aasta (ülikooli üldõppe diplom, loodusteadused eriala)

Töökohad ja ametid:

2018–…	Tartu Ülikool, Tartu Ülikooli genoomika instituut, Eesti Biokeskus, teadur.
2009–2017	Tartu Ülikool, Eesti Biokeskus, Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituut, teadur.
2003–2009	Tartu Ülikool, Eesti Biokeskus, Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituut, osalise tööajaga teadur.
2001–2003	Tartu Ülikool, Eesti Biokeskus, Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituut, külalisüliõpilane.
1999–2001	INSERM U463, Nantes, Prantsusmaa, laborant
Summer 1999	Bioloogiliste ja biokeemiliste analüüside labor, Quimper'i Laennec haigla, Prantsusmaa.

Kursused:

2–4 Aprill, 2019	CodeRefinery seminar, "teaching researchers in sustainable software development", Tartu, Eesti
27–28 August, 2019	ELIXIR Python'i kursus, Tartu, Eesti
21–22 Mai, 2015	ELIXIR R'i algkursus, Tartu, Eesti
6–7 Juuni, 13–14, 2013	Andmete analüüs Python'iga, Tartu, Eesti
22–26 August, 2013	OpenGENE project's workshop "GWAS: From Genotyping to Sequencing", Tartu, Eesti
13–16 Mai, 2007	2nd Paris Workshop on Molecular and Statistical Genomic Epidemiology, Paris, Prantsusmaa

Teadupublikatsioonid:
loetletud inglisekeelse CV rubriigis "Publications"

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

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

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

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

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

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

104. **Andres Tover**. Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 126 p.
105. **Helen Udras**. Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005, 100 p.
106. **Ave Suija**. Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005, 162 p.
107. **Piret Lõhmus**. Forest lichens and their substrata in Estonia. Tartu, 2005, 162 p.
108. **Inga Lips**. Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005, 156 p.
109. **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.
354. **Monika Karmin.** Perspectives from human Y chromosome – phylogeny, population dynamics and founder events. Tartu, 2018, 168 p.
355. **Maris Alver.** Value of genomics for atherosclerotic cardiovascular disease risk prediction. Tartu, 2019, 148 p.
356. **Lehti Saag**. The prehistory of Estonia from a genetic perspective: new insights from ancient DNA. Tartu, 2019, 171 p.
357. **Mari-Liis Viljur**. Local and landscape effects on butterfly assemblages in managed forests. Tartu, 2019, 115 p.
358. **Ivan Kisly.** The pleiotropic functions of ribosomal proteins eL19 and eL24 in the budding yeast ribosome. Tartu, 2019, 170 p.
359. **Mikk Puustusmaa**. On the origin of papillomavirus proteins. Tartu, 2019, 152 p.

360. **Anneliis Peterson**. Benthic biodiversity in the north-eastern Baltic Sea: mapping methods, spatial patterns, and relations to environmental gradients. Tartu, 2019, 159 p.