DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
386

BURAK YELMEN

Characterization of ancient Eurasian influences within modern human genomes

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
386

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
386

BURAK YELMEN

Characterization of ancient Eurasian influences within modern human genomes

UNIVERSITY OF TARTU
Press

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

Dissertation was accepted for the commencement of the degree of Doctor of Philosophy in Gene Technology on June 15, 2021 by the Council of the Institute of Molecular and Cell Biology and Institute of Genomics, Faculty of Science and Technology, University of Tartu.

Supervisors: Luca Pagani, PhD; Associate Professor, Department of Biology, University of Padova, Italy; Senior Research Fellow of Population Genetics, Institute of Genomics, University of Tartu, Estonia

Mait Metspalu, PhD; Director and Professor of Evolutionary Genomics, Institute of Genomics, University of Tartu, Estonia

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

Opponent: Mehmet Somel, PhD; Associate Professor, Department of Biological Sciences, Middle East Technical University, Turkey

Commencement: Room No 105, 23B Riia St, Tartu, on August 24, 2021, at 10:15.

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

ISSN 1024-6479
ISBN 978-9949-03-654-7 (print)
ISBN 978-9949-03-655-4 (pdf)

Copyright: Burak Yelmen, 2021

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Figures

Tables

LIST OF ORIGINAL PUBLICATIONS

REF I
Yelmen, B., Mondal, M., Marnetto, D., Pathak, A.K., Montinaro, F., Gallego Romero, I., Kivisild, T., Metspalu, M. and Pagani, L., 2019. Ancestry-specific analyses reveal differential demographic histories and opposite selective pressures in modern South Asian populations. *Molecular Biology and Evolution*.

REF II
Yelmen, B., Marnetto, D., Molinaro, L., Flores, R., Mondal, M. and Pagani, L., 2021. Improving selection detection with population branch statistic on admixed populations. *Genome Biology and Evolution*.

REF III
Molinaro, L., Montinaro, F., **Yelmen, B.**, Marnetto, D., Behar, D.M., Kivisild, T. and Pagani, L., 2019. West Asian sources of the Eurasian component in Ethiopians: a reassessment. *Scientific Reports.*

Author's contributions:

REF I: Study design, data analysis, writing the manuscript.
REF II: Study design, data analysis, writing the manuscript.
REF III: Data analysis (local ancestry deconvolution).

ABBREVIATIONS

aDNA	ancient DNA
AF	African
AG	Artificial Genome
AHFD	Ancestral Haplotype Frequency Difference
ANI	Ancestral North Indian
ASI	Ancestral South Indian
ARB	Ancestral Random Breeder
CEU	Utah residents with ancestry from northern and western Europe
EHH	Extended Haplotype Homozygosity
EMBA	Early-Middle Bronze Age
FDR	False Discovery Rate
GAN	Generative Neural Network
GIH	Gujarati Indian in Houston, TX
HMM	Hidden Markov Model
ITU	Indian Telugu in the UK
MLBA	Middle and Late Bronze Age
N	West Eurasian component of contemporary South Asian genomes
NAF	Non-African
PBS	Population Branch Statistic
PCA	Principal Component Analysis
RBM	Restricted Boltzmann Machine
S	South Asian component of contemporary South Asian genomes
TPR	True Positive Rate
TSI	Toscani in Italy
UMAP	Uniform Manifold Approximation and Projection
XP-EHH	Cross-population Extended Haplotype Homozygosity

1. INTRODUCTION

Genomic composition of present-day human populations has been shaped by past demographic events such as migrations, bottlenecks, founder events and admixing of different groups. Studying these processes and understanding the history of our species, although still challenging, have become easier thanks to the advancements in sequencing technologies and ancient DNA (aDNA) applications along with new data analysis and modelling methods. Although aDNA is now a crucial tool for this endeavour, its availability is limited due to different geographical locations having varying conditions for long term survival of DNA causing both quality and quantity issues in many cases. DNA degradation occurs due to chemical processes related to these conditions and it is more prevalent in older samples. In addition, warmer climates usually prevent aDNA preservation entirely. Therefore, modern human genomes are still widely studied to understand our past as they bear signals of these ancient events.

The evolution of phenotypic features termed 'anatomically modern' extends in African fossil record back to at least 300,000 years before present (Hublin et al. 2017). Most genetic variation outside Africa derives from the recent "out of Africa" event, with two possible major routes of dispersal into Eurasia either through Sinai peninsula to the Levant (the 'Northern Route') and or through Bab el Mandeb strait of the Red Sea to Arabia (the 'Southern Route') 50,000–100,000 years ago (Derricourt 2005; Armitage et al. 2011). There is still no full consensus yet on these routes on the basis of current archeological and genetic evidence (Reyes-Centeno et al. 2015). After the out of Africa event, different population splits accompanied by evolutionary forces such as genetic drift and natural selection contributed to the diversity of human genomes worldwide. A defining period for the establishment of the present-day gene pools in certain continental regions such as Europe, South Asia and Africa was the migration and mixing of human populations that happened after the Neolithic. Populations settled in the Near East and Eurasian steppe admixed with autochthonous populations all over West and South Eurasia and the long wave of these migrations reached as far as East and South Africa.

The main aim of the thesis is to provide insights into the ancient Eurasian components of the contemporary human populations with a specific focus on South Asian and Ethiopian populations utilizing modern human genomes. Although aDNA has already proved to be valuable for this purpose, modern human genomes can also be seen as a data source made up of a combination of ancient layers which can be 'excavated'. In this light, South Asian genomes have been demonstrated to be composed of South Asian (S) and West Eurasian (N) components admixed in the last 10,000 years (Reich et al. 2009). Ethiopians have been characterized as having multiple distinct African ancestry components and a Eurasian component derived from the Near East 3,000 years ago. This thesis includes a closer look at the demography of these two major components of South Asian populations (REF I), methodological approaches to improve selection detection in admixed populations and possible selection signals from South Asian genomes (REF I, II), and insight into the origins of Eurasian component of Ethiopians (REF III) with the help of local ancestry inference methods.

2. LITERATURE OVERVIEW

2.1. Post-Neolithic admixtures

After the out of Africa event, founder populations were settled in Eurasia followed by the Neolithic transition 10,000–15,000 years ago which marked the beginning of human agricultural practice in the Near East (Bocquet-Appel 2011). Major migrations from the Near East to Europe, Asia and Africa coupled with the spread of agriculture in the Neolithic and post Neolithic era molded the mosaic of the contemporary human genomes (**Figure 1**) (Lazaridis et al. 2014; Lazaridis et al. 2016). Among the affected populations were South Asian populations which were admixed with farmers from Near East and Eurasian steppe (Reich et al. 2009; Narasimhan et al. 2019), and Ethiopians which were admixed with Eurasian sources possibly from the Levant (Pagani et al. 2012; Hodgson et al. 2014; Lazaridis et al. 2016).

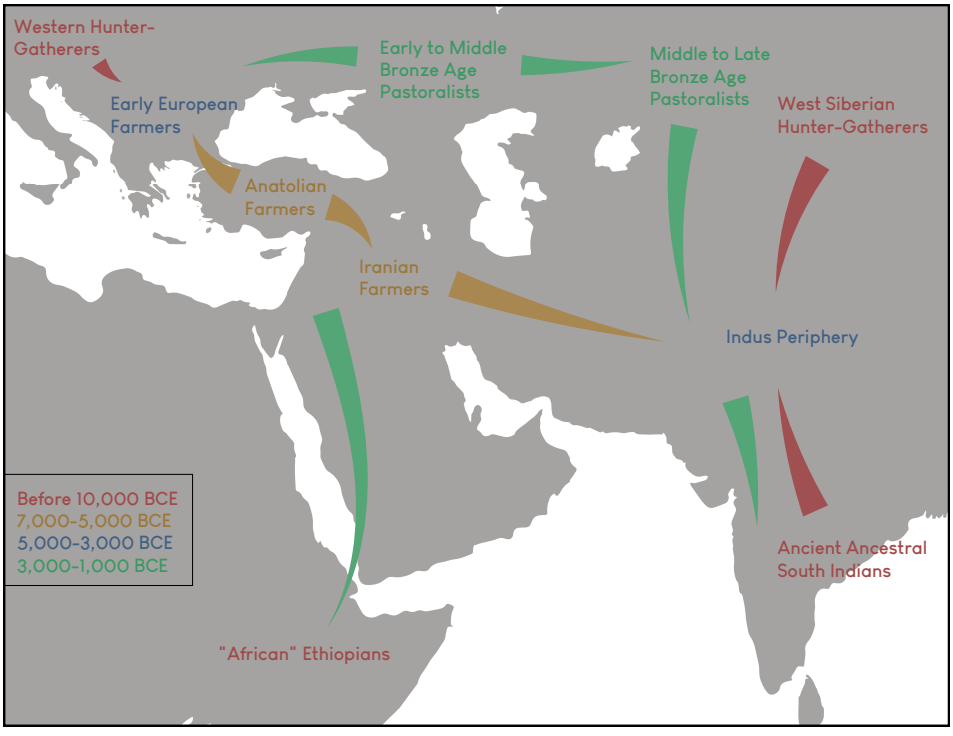

Figure 1. A simplistic model of post-Neolithic migrations. Arrows indicate direction of gene flow. Information is based on Pagani et al. 2012, Lazaridis et al. 2014, Metspalu et al. 2018, Narasimhan et al. 2019. Colors solely indicate time periods and are not linked to inferred genetic components.

It is possible to detect and dissect these contributions in modern human genomes and study these components which were brought from elsewhere to shed light into past demographic events (Pagani et al. 2015). This approach is especially important when aDNA availability is limited. It is also viable to interpret natural selection acting on these components both to discover potential signals and to distinguish the source of adaptive variations in human genomes (Hodgson et al. 2014).

2.2. Complex demography of South Asia

After dispersal to Eurasia from Africa, South Asia was occupied by anatomically modern humans at least for 36,000–38,000 years based on fossil evidence while archaeological findings suggest 50,000 years or more (Mellars et al. 2013; Bae et al. 2017). Studies have shown that modern South Asian genomes are composed of autochthonous South Asian (S) and West Eurasian components (N), with proportions varying in the Indian cline defined by a diverse population structure (Reich et al. 2009; Chaubey et al. 2011; Metspalu et al. 2011) (**Figure 2**). These two components represent one of the deepest splits after the out of Africa event with an at least 40,000-year difference. A recent study based on available aDNA from the region detailed this composition in more detail, specifically defining three major ancestral sources as autochthonous South Asian (South Asian hunter-gatherer), Indus Periphery and Steppe Middle and Late Bronze Age (Steppe MLBA) (Narasimhan et al. 2019). Yet there are still no ancient genomes in unadmixed form available to represent the ancient autochthonous South Asians. This is possibly due to the hot and humid conditions which prevail in this geographic region being unfavorable for skeletal and DNA preservation. This hurdle combined with the complex demographic history and diverse selective forces in South Asia makes disentangling the compound of genetic ancestries in the South Asian genomes a challenging task.

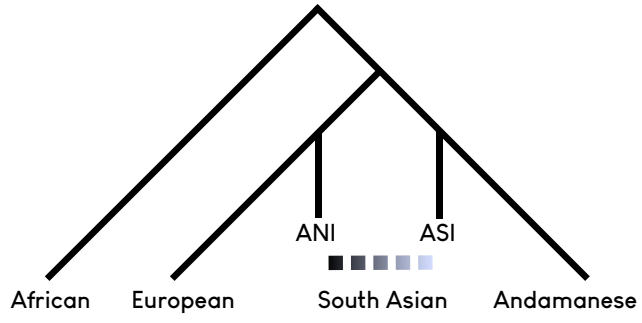

Figure 2. Components of South Asian populations. Model of South Asian genomes as a mixture of "Ancestral North Indian" (ANI) and "Ancestral South Indian" (ASI) components demonstrated in the context of a phylogenetic tree. Dashed color gradient represents varying proportions of ANI and ASI ancestries. Branch lengths are not proportional to split times. Information based on Reich et al. 2009, Chaubey et al. 2011, Metspalu et al. 2011.

2.3. Genomic mosaic of Ethiopians

Ethiopia and the Horn of Africa, exhibiting great diversity in culture and language among its inhabitants, lies in Africa closest to Arabian peninsula as a Southern bridge point out of Africa (Pagani et al. 2012). Based on archeological and genetic findings, it is considered to be a strong candidate route during human expansion from Africa to Eurasia 50,000 to 100,000 years ago (Armitage et al. 2011; Pagani et al. 2015). Aside from the archeological and linguistic clues related to migration patterns which formed the contemporary Ethiopian populations, recent genetic studies suggest precolonial gene flow both from within Africa and from Eurasia dating back to 3,000 years ago (Pagani et al. 2012; Pickrell et al. 2014; Lazaridis et al. 2016), which is also supported by linguistic evidence (Kitchen et al. 2009). Although a Levantine based source is suggested for this recent Eurasian admixture, it is difficult to pinpoint the exact source given the subtle differences in genetic variation for the possible candidates.

2.4. Dissecting Eurasian component in modern populations

2.4.1. Local ancestry inference

Genomes can be seen as a blend of different ancestral contributions; these ancestries can be identified and studied independently with different methods. Global ancestry inference refers to estimating proportions of different ancestries in admixed individuals. A notable approach used for this purpose is population structure inference tool STRUCTURE which uses a Bayesian algorithm to define K number of clusters (populations) and probabilistically assigns individuals to these clusters (Pritchard et al. 2000). ADMIXTURE is a similar method later implemented with the same basis but relies on maximum-likelihood instead of sampling the posterior distribution which allows it to be run on datasets with hundreds of thousands of markers (Alexander et al. 2009). Local ancestry inference, on the other hand, is the estimation of ancestry for genomic segments in admixed genomes based on reference populations. Several algorithms have been proposed for this task. Some notable examples are PCAdmix which uses a principal component analysis (PCA) based approach (Brisbin et al. 2012), ELAI and MOSAIC which both use a two-layer hidden Markov model (HMM) (Guan 2014; Salter-Townshend and Myers 2019) and RFMix which uses a random forest trained by reference panels (Maples et al. 2013) (**Table 1**).

Table 1. Comparison of four local ancestry inference methods.

Software	Algorithm	Phasing	Genetic map	Ancestral sources
PCAdmix	PCA	Required	Required	Predefined
ELAI	two layer HMM	Not required	Not required	Predefined
MOSAIC	two layer HMM	Required	Required	Not predefined
RFMix	Random forest	Required	Required	Predefined

More in detail, PCA is used in the PCAdmix method to assess the contribution of each SNP to the ancestry classification of a genomic region. Obtained principal components (PCs) are used as weights to get the weighted average of predefined windows of SNPs, and these averaged scores are fed to a hidden Markov model (HMM) as the observed values to obtain posterior probabilities for the ancestry of each window (Brisbin et al. 2012). ELAI, on the other hand, uses source populations to model two linkage disequilibrium levels of "within" and "between" haplotype groups thanks to the two-layer HMM model. In these two layers, clusters are labelled to demonstrate ancestral alleles and multiple clusters under the same label over neighbouring markers demonstrate ancestral haplotypes. The upper layer clusters represent populations, or structure closer to the root of a coalescent tree, and the lower layer clusters represent within population haplotypes, or structure closer to the tip of a coalescent tree. Aside from the benefit of fine-tuned local ancestry inference, ELAI can also be applied to unphased diploid data unlike many other methods, eliminating the need for phasing the genomic data and thus, phase uncertainty (Guan 2014).

After ancestral assignment is achieved with any given method, it is possible to study a particular ancestry by creating "masked" genomes which consist only of genomic chunks belonging to the target ancestry/ancestries.

2.4.2. Allele frequency-based analyses

Once the ancestral components are extracted, they can be characterized from a demographic perspective. Most analyses regarding population structure are based on differences in variant frequencies between different groups. Based on the allele data, dimension reduction techniques such as principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) are commonly used in population genetics as initial screening methods to assess population structure (Cavalli-Sforza et al. 1994; McInnes et al. 2018; Diaz-Papkovich et al. 2019). As genomic data is difficult to summarize due to its multidimensional characteristic, these feature extraction approaches are useful for both interpretation and visualisation of variation among human populations. Additionally, softwares such as ADMIXTURE are used for clustering to estimate individual ancestries with a maximum-likelihood approach based on SNP data (Alexander et al. 2009). F-statistics, based on shared genetic drift, have also been used widely to test hypotheses about admixture between populations as shared drift between different

populations is a sign of shared evolutionary history (Reich et al. 2009; Patterson et al. 2012). In further detail, f-statistics compute allele frequency correlations among two, three or four populations to provide insight into population splits and admixture events. f4-statistic (or D-statistic) measures average correlation of allele frequency differences between population A and B, and population C and D denoted with the commonly used notation f4(A, B; C, D). If true phylogeny for these populations is as ((A,B),(C,D)), then we expect f4(A, B; C, D) to be zero and f4(A, C; B, D) and f4(A, D; B, C) to be positive. In the same spirit, f2-statistic can be defined as f2(A, B) = f4(A, B; A, B) and f3-statistic can be defined as f3(A; B, C) = f4(A, B; A, C). Additionally, if population A is an admixture of ancestral A_1 and A_2 populations with proportions α and (1-α) respectively, then f4(A, B; C ,D) is expected to be αf4(A_1, B; C, D) + (1-α)f4(A_2, B; C, D) (Lipson 2020).

2.4.3. Detecting natural selection

As human populations dispersed around the world, they were subject to different environmental conditions, diets and pathogens. Different selective forces have acted on human phenotypes in different environments which consequently contributed to the genomic diversity we observe. Studying these adaptive changes is crucial for better understanding of evolutionary mechanisms and advancing medicine. In terms of genomics, an allele whose frequency is changing faster or longer towards the same direction is considered to be under selection. To detect this in the human genome, different methods have been used, which are frequency-based, linkage disequilibrium-based, population differentiation-based, composite and more recently machine learning-based methods (Vitti et al. 2013; Sheehan and Song 2016). A notable example for frequency-based methods is population branch statistic (PBS), which compares pairwise FST values between three populations to detect whether a locus has an unusual amount of genetic drift in the target population compared to the other two. PBS has been shown via simulation studies to have strong power to detect recent natural selection (Yi et al. 2010).

Selective sweeps in the genome occur when a beneficial mutation increases in frequency and becomes fixed, causing the genetic variation to be reduced in the genetic neighbourhood of the mutation due to linked alleles. The increase of the frequencies of alleles linked to the beneficial allele is called genetic hitchhiking. One way to detect selective sweeps is linkage-disequilibrium based approaches such as cross-population extended haplotype homozygosity (XP-EHH) (Sabeti et al. 2007). The EHH (extended haplotype homozygosity) statistic part of XP-EHH computes the probability that two extended haplotypes around a locus are the same given that both have the same allele, to quantify reduction in haplotype diversity. In other words, it provides a homozygosity metric for extended haplotypes instead of alleles. XP-EHH compares the EHH between two populations at the same SNP (Wagh et al. 2012). Therefore, it can detect selective sweeps by detecting alleles which have risen to high frequency or fixation in one

population, yet observed as polymorphic in the other. Although commonly used in selection studies, efficacy of these methods in admixed populations is understudied and it remains a challenge to identify whether a detected signal is due to selection after or before the admixture event (Huerta-Sánchez et al. 2013; Hodgson, Pickrell, et al. 2014; Huerta-Sánchez et al. 2014).

3. AIMS OF THE STUDY

The aim of this thesis is to characterize demographic and adaptive features of ancient Eurasian components in modern human populations with a particular focus on South Asia and Ethiopia with the help of available genomic data and methods for local ancestry inference in admixed individuals. More specific goals of each publication are as follows:

First study (REF I) investigates the past demographic events in South Asia by creating surrogates of ancient South Asian and West Eurasian components which constitute the contemporary genomes in the region. It aims to provide insights on how the diverse variation in South Asian genomes came to be and to explore possible recent selection signals by detecting imbalances in the reconstructed components.

Second study (REF II) expands on the idea of the first one and investigates whether ancient components carved out of modern genomes can be used to improve detecting selection in admixed individuals. It provides comparative analysis on how reliable selection signals in admixed genomes are and additionally, tries to detect first possible selection signals in the South Asian component of modern South Asians.

Third study (REF III), similarly to the first two, utilizes local ancestry inference and allele sharing methods to study the Eurasian component in modern Ethiopian genomes. It aims to pinpoint the source of the Eurasian admixture with the help of available modern and ancient genomes.

4. MATERIALS AND METHODS

All data used in the three publications have been obtained from the literature. No new sample collection or genetic data generation was undertaken. Samples used in this thesis can be seen in the Appendix section. Below is the basic workflow (**Figure 3)** followed by a summary of methods used, detailed information can be found in the original manuscripts.

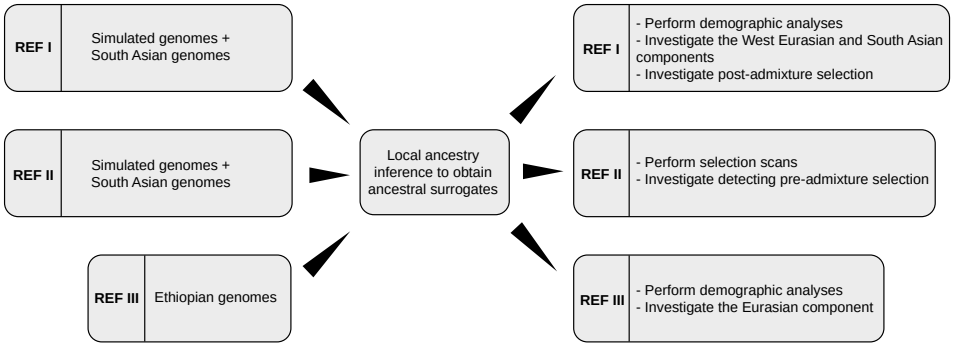

Figure 3. Basic workflow of the three studies.

4.1. REF I

In this study, we first simulated proxies for European, Asian and African populations via ms software (Hudson 2002) with model parameters from (Jouganous et al. 2017) with the following command line with seeds (x x x) ranging from 1 to 22 in the form of chromosomes:

ms 80 1 -t 6493.248 -r 5861.96 10000000 -I 4 20 20 20 20 -n 1 8.977 -n 2 4.175 -n 3 2.104 -n 4 5.457 -g 1 267 -g 2 172.35 -g 4 235.5075 -m 1 2 1.889 -m 2 1 1.889 -m 1 3 0.216 -m 3 1 0.216 -m 2 3 0.496 -m 3 2 0.496 -es 0.0038236 4 0.70 -ej 0.007647201 4 2 -ej 0.007647201 5 1 -ej 0.034 1 2 -em0.0956 2 3 7.124 -em 0.0956 3 2 7.124 -ej 0.0956 2 3 -en 0.239 3 1 -eA 0.006882481 4 10 -eA 0.006882481 5 10 -p 15 -seeds x x x

An Indian-like population was generated as an admixture of ancient European (N) and ancient Asian (S) simulated populations. We performed local ancestry deconvolution on the Indian-like proxy via PCAdmix (Brisbin et al. 2012) with window size 10 and default parameters and obtained masked genomes (MASK_S, European chunks masked out and MASK_N, Asian chunks masked out) based on ancestry assignment of genomic chunks to assess if we can retrieve proxies for the two ancient sources. We set 95% threshold for the PCAdmix fbk output as a confidence limit below which the ancestry assignment was deemed as not

reliable, and we used only chunks assigned with a probability above this threshold while creating the MASK haplotypes. Additionally, we generated ancestral random breeders (ARBs) by replenishing the masked haplotypes (MASK_S and MASK_N) from other masked donors within the same population to acquire full genomes for the ancient proxies. ARBs, therefore, can be seen as a set of random breeders of ancestral source populations with genomes without gaps. We tested our approach with PCA and ADMIXTURE on these simulated data to demonstrate that proxies for the ancient components can be retrieved as MASK and ARB haplotypes. We applied our method on real genomes using SNP array data of 565 contemporary samples from South Asia (Li et al. 2008; Altshuler et al. 2010; Metspalu et al. 2011; Basu et al. 2016; Pathak et al. 2018) along with additional 404 contemporary and 360 ancient samples (Behar et al. 2010; Yunusbayev et al. 2012; Behar et al. 2013; Haak et al. 2015; Yunusbayev et al. 2015; Mörseburg et al. 2016; Narasimhan et al. 2019) and 1000 Genomes data (1000 Genomes Project Consortium et al. 2015). We performed local ancestry deconvolution on admixed South Asian genomes via PCAdmix (Brisbin et al. 2012) with French and Paniya as proxies for West Eurasian (N) and South Asian (S) components of South Asian genomes, respectively, to obtain MASK and ARB haplotypes. Since ARB production may introduce spurious allele frequency shifts, we used ARBs only for PCA and ADMIXTURE analyses as we demonstrated them to be viable for these methods via simulations. For frequency-based methods such as f3 and f4-stats, qpAdm and qpGraph, we used AdmixTools 4.1 software (Patterson et al. 2012) utilizing MASK haplotypes. We additionally investigated post-admixture selection on South Asian genomes. For this purpose, we defined the metric Ancestral Haplotype Frequency Difference (AHFD) which evaluates deviations from N and S ancestry on genomic regions based on local ancestry assignment since high deviations from the ancestry proportion would point towards recent selection after admixture.

4.2. REF II

In this study, we used a subset of the data from REF I to first simulate 500 samples as an admixture of real French and Han genomes using the forward-simulator software admix-simu (Williams 2016). We performed local ancestry inference in the same manner as REF I via PCAdmix (Brisbin et al. 2012) and additionally using ELAI (Guan 2014) with German and Japanese as proxies for N and S ancestries to obtain MASK and ARB haplotypes. We performed population branch statistic (PBS) analyses on MASK_N, naive (population simulated as an admixture of French and Han) and French (source) populations with outgroups Yoruba and Japanese [PBS(N, Japanese, Yoruba)]. We also performed PBS on MASK_S, naive and Han (source) populations with outgroups Yoruba and German [PBS(S, German, Yoruba)]. Cross-population extended haplotype homozygosity (XP-EHH) was performed on ARB_N, naive and French (source) populations with Japanese outgroup. XP-EHH was also performed on ARB_S, naive and Han

(source) populations with German outgroup. We used ARBs for XP-EHH as the method requires full haplotypes. To compare the results, we used Spearman's correlation coefficient and bootstrapping to acquire confidence intervals. We additionally performed 50 kb window-based comparisons for the PBS results of MASK, naive and source populations by calculating true positive rates (TPR) and false discovery rates (FDR). After demonstrating improvement in detecting pre-admixture selection with PBS on simulated genomes, we applied this approach on real South Asian genomes. First we performed PCAdmix on real South Asian genomes using French and Paniya as proxies for N and S ancestries, respectively, to acquire MASK_S and MASK_N haplotypes. We performed PBS on retrieved MASK_S haplotypes to report the first possible selection signals on the S component of contemporary South Asian genomes.

4.3. REF III

In this study, we used 1704 modern and ancient samples including 5 contemporary Ethiopian populations (Amhara, Gumuz, Oromo, Ethiopian Somali and Wolayta) (Pagani et al. 2015) and a number of reference populations (Li et al. 2008; Behar et al. 2010; Pagani et al. 2012; Patterson et al. 2012; Pickrell et al. 2012; Behar et al. 2013; Lazaridis et al. 2014; Raghavan et al. 2014; 1000 Genomes Project Consortium et al. 2015; Mathieson et al. 2015; Allentoft et al. 2015; Gallego Llorente et al. 2015; Jones et al. 2015; Pagani et al. 2015; Broushaki et al. 2016; Fu et al. 2016; Hofmanová et al. 2016; Kılınç et al. 2016; Lazaridis et al. 2016; Omrak et al. 2016; Haber et al. 2017; Lazaridis et al. 2017; Lipson et al. 2017; Skoglund et al. 2017; Harney et al. 2018; Feldman et al. 2019). We used MALDER software for dating the admixture event and determining the number of waves of admixture (Pickrell et al. 2014). We performed local ancestry inference via PCAdmix as in REF I with CEU (Utah residents with ancestry from northern and western Europe) and Gumuz (Ethiopian population with the lowest Eurasian component content) populations as proxies for Non-African (NAF) and African (AF) components, respectively, to retrieve masked NAF (AF assigned chunks masked out) and AF (NAF assigned chunks masked out) haplotypes.

In addition, we classified ancestry assigned windows as high confidence non African (NAF) (assigned to non African with >90% confidence), low confidence non African (X) (assigned to non African with 51–90% confidence), high confidence African (AF) (assigned to African with >90% confidence) and low confidence African (Y) (assigned to African with 51–90% confidence) based on PCAdmix fbk threshold to account for possible biases introduced via local ancestry deconvolution and demonstrated that X and Y results do not deviate much from the analyses run using only the genetic components inferred with high confidence. To put it another way, X and Y components which include ancestral assignments with low confidence (unlike NAF and AF) did not have additional genomic regions which can alter the results.

We utilized PCA for initial screening and ADMIXTURE for supervised clustering analysis. Additionally, we performed outgroup f3 statistic analysis via POPSTATS software (Skoglund et al. 2015) and f4 statistic analysis along with qpWave and qpAdm via Admixtools 4.1 (Patterson et al. 2012) to assess demographic model and admixture events. We also replicated the local ancestry deconvolution procedure using ELAI (Guan 2014) and retained the same f4 results.

5. RESULTS AND DISCUSSION

This section includes the summary of the three scientific publications with novel discoveries. Original articles can be examined for detailed information.

5.1. Investigating demography and post-admixture selection in South Asian genomes (REF I)

5.1.1. Demographic assessment of genomic components of South Asia

After demonstrating that proxies for ancient components can be recovered from admixed genomes via simulations both in the form of masked (MASK) and complete (ARB) haplotypes (see Materials and Methods), we performed PCA with ARB and contemporary real genomes as an initial screening (**Figure 4**). On PCA space of the first two components, reconstructed ARB_N haplotypes are placed between Near East and Pakistan samples and ARB_S haplotypes are placed further down the Indian cline as expected. Separation of N and S clusters is observed both on the first PC, which differentiates West and East Eurasian populations, and on the second PC, which differentiates East Asia and Andaman following the placement of real genomes from which they were constructed.

To further investigate possible proxies for reconstructed haplotypes, we performed outgroup f3 analysis as f3(X, Y; Yoruba) with X being MASK_S and MASK_N populations and Y being all other populations including ancient samples from Narasimhan et al. (2019). Based on this, closest proxies to MASK_S were other MASK_S populations followed by Southern Indian populations (such as Paniya, source used for ancestry deconvolution, and Irula) and Han and Onge. Closest ancient proxies were Saidu Sharif samples from Pakistan (500–300 BCE) and Indus diaspora samples from Iran (2550–2450 BCE). Similarly, closest proxies for MASK_N were mainly other MASK_N populations along with French (source used for ancestry deconvolution) but additionally ancient Eurasian samples such as Steppe_EMBA (Yamnaya Early-Middle Bronze Age).

To test the expected tree-like behaviour of the masked proxies, we calculated D-statistics in form of D(Gujaratis_N, French; Onge, Mbuti; Z=–0.827), D(Gujaratis_N, French; Han, Mbuti; Z=0.152), D(Irula_S, Onge; Steppe_EMBA, Mbuti; Z=0.270), D(Irula_S, Onge; French, Mbuti; Z=–2.357) using Gujaratis_N as a representative of MASK_N and Irula_S as a representative of MASK_S populations. The insignificant deviations from zero (absolute value of Z score < 3) demonstrate that the deconvolution approach produced relatively clean MASK haplotypes without admixture. To investigate the phylogenetic relationship between the S component, Onge and Han, we also obtained D(Irula_S, Han; Onge, Mbuti; Z=0.268), D(Irula_S, Onge; Han, Mbuti, Z=–0.068) and D(Han, Onge; Irula_S, Mbuti; Z=–0.348), which suggest that Irula_S, Onge and Han do not conform to a bifurcating tree as each forms an independent basal lineage, as

also seen from the PCA plot (**Figure 4**). On the other hand, qpAdm analysis showed relatively clean MASK_N haplotypes whereas MASK_S haplotypes had a varying amount of West Eurasian admixture. This might both be related to the deficiency in ancestry deconvolution method or to a Neolithic/pre-Neolithic West Eurasian gene flow to the S component.

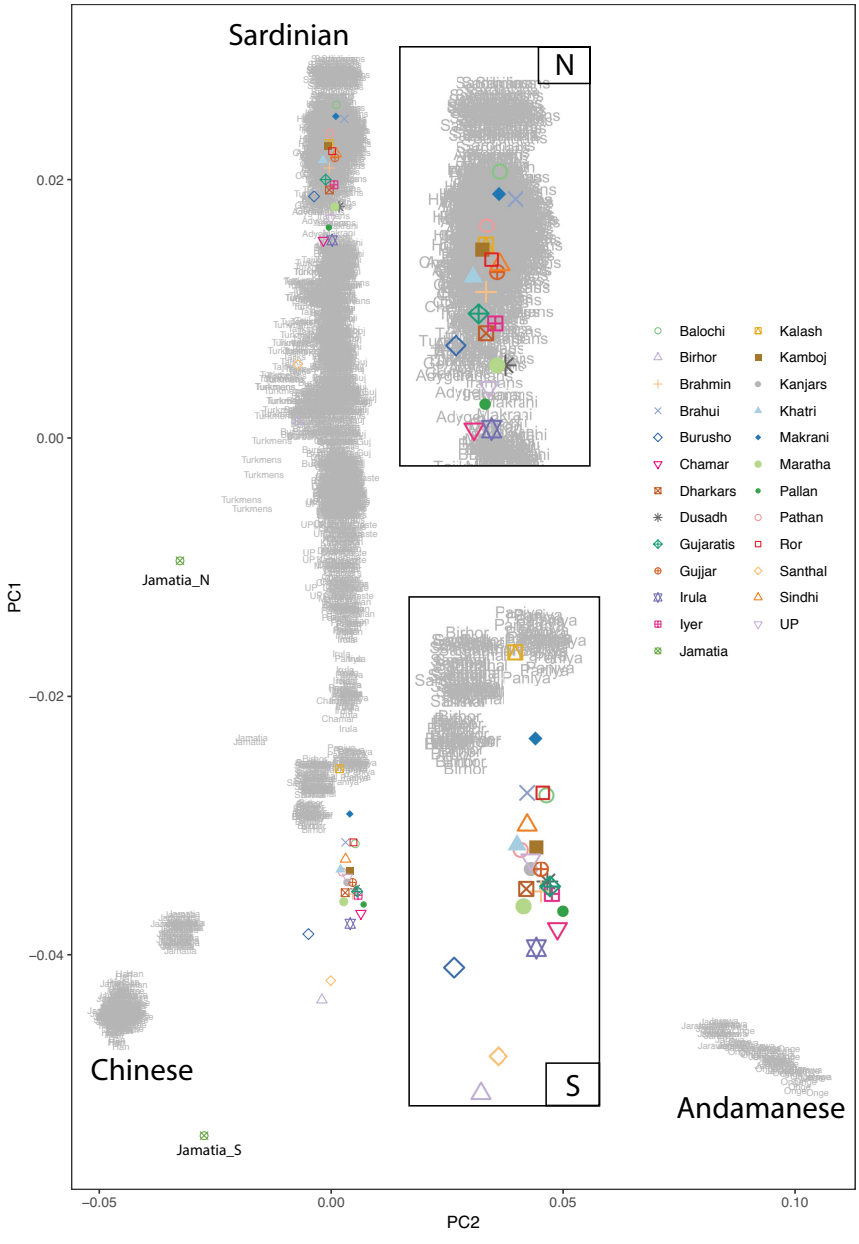

Figure 4. PCA of reconstructed South Asian components and contemporary populations. Each point is a haploid ARB (colored) or modern (grey) sample. Insets (N and S) zoom on ARB_N and ARB_S clusters respectively. Figure from Figure 1 (Yelmen et al. 2019), licenced under CC BY-NC 4.0.

We additionally tried to produce a demographic model of MASK haplotypes and possibly relevant populations using gpGraph to gain further grip on N and S ancestries. We were able to model a simpler tree structure for MASK_S haplotypes whereas we couldn't produce fitting models for MASK_N, although invoking multiple admixture events improved the fit. This suggests a relatively unadmixed S component and multiple waves of arrival for the N component in the region. Moreover, we were able to model both N and S components in a broader picture without any f2 and f4 outliers and demonstrated again S, East Asia and Andamanese as a trifurcation (**Figure 5**).

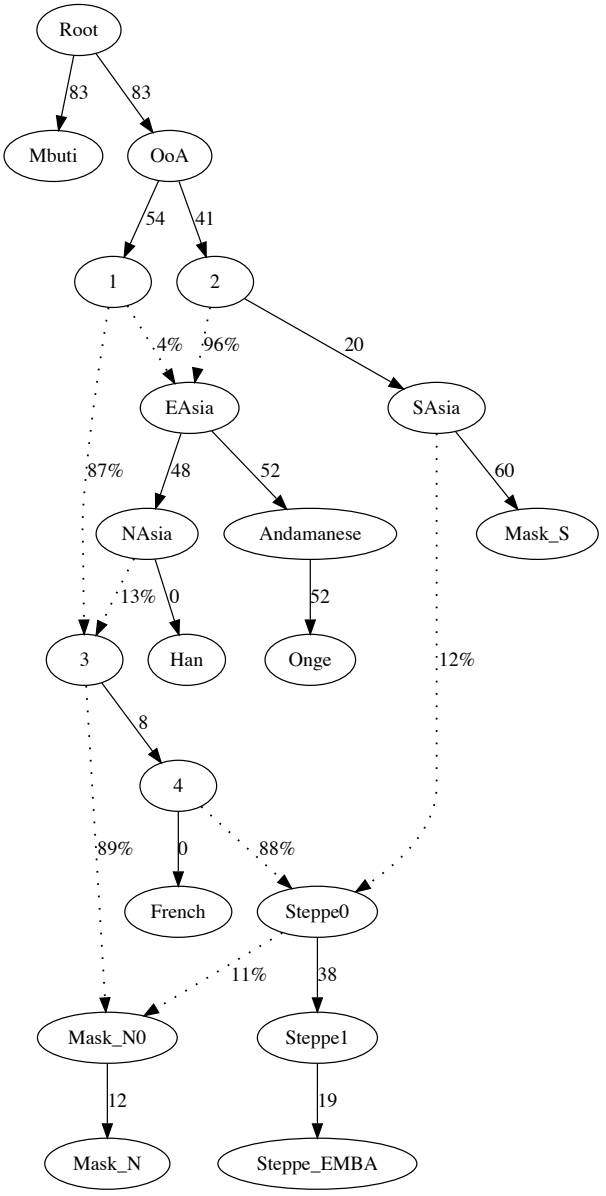

Figure 5. gpGraph model of N and S components. Other models with different Andamanese, Han and S splits yielded f4 outliers and poorer fit. Final score: 3061.001, degrees of freedom: 2, no f2 outliers, no f4 outliers, worst f-stat: 1.747. Figure from Figure 2 (Yelmen et al. 2019), licenced under CC BY-NC 4.0.

5.1.2. Differences in phenotype-informative SNPs between components

We collected a list of phenotype-informative SNPs from the literature (Mathieson et al. 2015; Gelabert et al. 2017; van de Loosdrecht et al. 2018) and identified differences in allele frequencies for these SNPs between MASK_N and MASK_S haplotypes to detect alleles which were brought to South Asia via the N ancestry. For this purpose, we followed the same local ancestry inference approach we used for the array data but instead utilized 1000 Genomes sequence data to increase SNP coverage. Deconvolution was applied to GIH population (Gujarati Indian in Houston, TX), with TSI population (Toscani in Italy) as a proxy for the N source and ITU (Indian Telugu in the UK) as a proxy for the S source. Instead of directly interpreting the distinction, we compared the allele frequency difference of the selected SNPs between GIH_N and GIH_S against source populations TSI and ITU, to take possible local ancestry deconvolution missassignments into account. We detected that the rs16891982 (G) allele, located on the *SLC45A2* gene and associated with skin and hair pigmentation in Europeans, is possibly linked with the arrival of N ancestry. Also lactase persistence allele rs4988235 (A) related to the *MCM6/LCT* gene seems to be linked to the N ancestry, but frequencies for this allele were very similar for ITU and TSI, suggesting this link might be an artifact of poor ancestry deconvolution. It is important to underline here that ITU is not the best option for the local ancestry deconvolution step due to high West Eurasian admixture, yet it was the optimal choice within available 1000 Genomes data.

5.1.3. Post-admixture selection

Following the Out-of-Africa dispersal and their initial departure, populations carrying the ancestral N and S components would have been exposed to different environmental conditions. When the N component arrived in South Asia, it likely underwent allele frequency shifts due to different selective pressures in this new environment. To investigate this, we defined a new metric, ancestral haplotype frequency difference (AHFD), which assesses local admixture imbalance between ancestral components. We examined the regions with the highest admixture imbalance and found excess N ancestry for *SETD5* (possibly involved in insulin sensitivity) (Palmer et al. 2015; Walford et al. 2016), *ZNF* (possibly related to colorectal cancer due to diet) (Figueiredo et al. 2014) and *HLA* (immunity connected) (Fairfax et al. 2012) genic regions. For the tests highlighting the excess of S ancestry, one key finding was the region containing *SLC45A2* gene which has been reported to be related to eye and skin pigmentation in West Eurasians (Mathieson et al. 2015; Crawford et al. 2017; Martin et al. 2017) (**Table 2**).

Table 2. Top five hits for N and S excess regions. Positive values are N excess whereas negative values are S excess. Table from Table 1 (Yelmen et al. 2019), licenced under CC BY-NC 4.0.

Com-ponent	Position (Chr:Start-End) (GRCh37)	Number of populations with significant AHFD values	Genes (±50 kb region)	Number of 10 SNP regions
N	3:9,363,925–9,595,374	22 (percentile = 99.9949)	*THUMPD3,* ***SETD5-AS1***	2
	6:84,399,772–85,572,756	21 (percentile = 99.9814)	*SNAP91, RIPPLY2, CYB5R4, MRAP2, CEP162, TBX18*	2
	6:30,079,993–30,257,693	21 (percentile = 99.9814)	*TRIM31, TRIM40, TRIM10, TRIM15, TRIM26,* ***HLA-L***	2
	14:97,636,701–97,715,909	19 (percentile = 99.9383)	*Intergenic*	1
	19:23,930,879–24,368,053	18 (percentile = 99.9195)	***ZNF681, ZNF726, ZNF254***	1
S	5:33,944,217–34,032,014	−21 (percentile = 0.0057)	*RXFP3,* ***SLC45A****, AMACR, C1QTNF3, ADAMTS12*	2
	20:652,097–694,894	−16 (percentile = 0.038)	*SRXN1, SCRT2, SLC52A3*	1
	8:116,208,407–116,308,464	−16 (percentile = 0.038)	*Intergenic*	1
	9:12,276,668–12,460,256	−15 (percentile = 0.0757)	*Intergenic*	2
	8:54,578,044–55,071,319	−14 (percentile = 0.1268)	*ATP6V1H, RGS20, TCEA1, LYPLA1, MRPL15*	1

5.2. Detecting selection in admixed populations (REF II)

After investigating post-admixture selection in South Asian populations, the next step would be to focus on selection before the admixture. However, research on selection detection in admixed populations is limited. Selection signals detected in an admixed population might be due to adaptation which happened in one of the ancestral groups, or signals present in the ancestral groups might cancel each other, rendering them undetectable in the admixed population. This study evaluates the effectiveness of the local ancestry deconvolution approach explained in REF I to detect selection in components of admixed genomes.

5.2.1. Forward simulation-based PBS comparison

After creating a French-Han admixed population with forward simulations (as a subtle mimicry of South Asian N and S admixture), we performed local ancestry inference as described in Materials and Methods which yielded MASK_N (S ancestry chunks masked out) and MASK_S (N ancestry chunks masked out) genomes. This allowed us to make comparisons between PBS results of MASK and naive (French-Han admixture) populations against source populations (French or Han). Overall SNP by SNP PBS comparison showed significantly higher correlation for MASK and source populations compared to naive and source populations. As PBS scores below a certain threshold can be disregarded as definitely uninformative, we additionally made comparisons based on positions which have PBS values above 99% threshold in source populations. These loci showed significantly higher correlation for MASK_N_ELAI and MASK_S_ELAI (masked populations generated based on ELAI) whereas MASK_N_PCAdmix and MASK_S_PCAdmix (masked populations generated based on PCAdmix) performed poorly (**Figure 6**). This difference is due to false negatives, meaning that PBS score is high for a given position for the source population but low for the MASK.

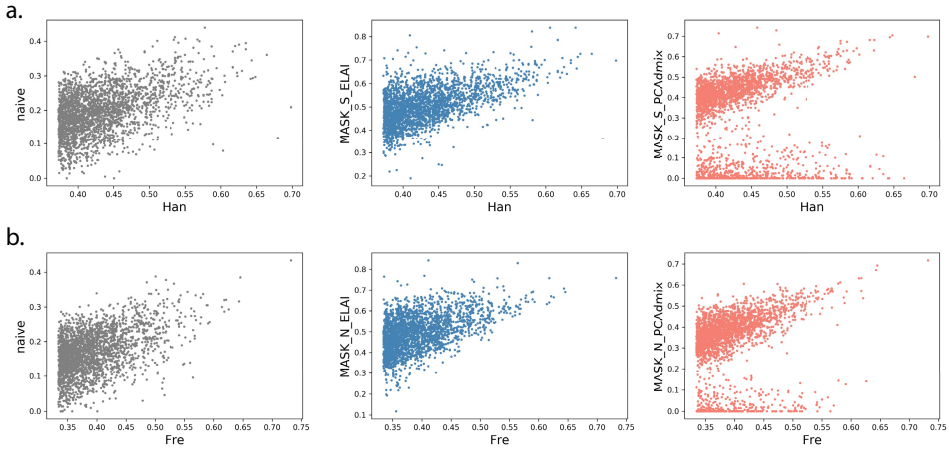

Figure 6. Comparing PBS values for each position with PBS score above 99% threshold. SNPs above threshold were selected based on Han and French source populations. **a)** Han vs naive (correlation coefficient: 0.409, 95% confidence interval: 0.377–0.441, p-value <2.2e16, n=2529), Han vs MASK_S_ELAI (correlation coefficient: 0.510, 95% confidence interval: 0.479–0.542, p-value <2.2e16, n=2529), Han versus MASK_S_PCAdmix (correlation coefficient: 0.210, 95% confidence interval: 0.169–0.252, p-value <2.2e16, n=2529). **b)** French vs naive (correlation coefficient: 0.400, 95% confidence interval: 0.367–0.436, p-value <2.2e16, n=2530), French versus MASK_N_ELAI (correlation coefficient: 0.452, 95% confidence interval: 0.422–0.483, p-value <2.2e16, n=2530), French versus MASK_N_PCAdmix (correlation coefficient: 0.394, 95% confidence interval: 0.354–0.432, p-value <2.2e16, n=2530). Figure from Figure 1 (Yelmen et al. 2021), licenced under CC BY-NC 4.0.

5.2.2. Forward simulation-based XP-EHH comparison

For XP-EHH, we generated ancestral random breeders (ARBs) as described in REF I as XP-EHH cannot be performed with MASK haplotypes due to the introduced missing portions. Similarly to the PBS case, we compared naive and ARB XP-EHH results against the source populations (French and Han). Comparisons showed that correlation between naive and source populations is higher than the correlation between ARB and source populations, suggesting utilizing ARBs does not improve XP-EHH results possibly due to ARB production procedure disrupting haplotypic structure. Additionally, although having better XP-EHH results than ARBs, naive populations also failed in having good correlation with the source populations. In other words, "true" selection signals that were present in the ancestral populations were mostly not caught either by ARB or naive populations. In addition, we detected false positive signals for both naive and ARB groups.

Remarkably, we could identify genomic regions with a signal of positive selection (>2 XP-EHH score) for the admixed population, although we did not simulate any selection, which suggests that signals detected in admixed populations via XP-EHH might not be related to selective pressures on that population but rather be due to one of the ancestral populations. More intriguingly, some signals we detected in the admixed population were not detected in the source populations which is possibly due to admixture events causing pseudo XP-EHH signals. Therefore, we urge future studies to be cautious while performing XP-EHH on admixed populations.

5.2.3. A case application on real South Asian genomes

Population genetics of South Asian populations is rather intricate due to multiple waves of Eurasian gene flow to the region (Narasimhan et al. 2019). Although recent studies based on aDNA have helped us to better understand the overall picture, there is still a lack of aDNA for the autochthonous South Asian component. Therefore, we applied our local ancestry inference-based method on real South Asian genomes to detect possible selection signals on the South Asian component. For this purpose, we initially created MASK_N and MASK_S haplotypes following the same procedure as simulations and performed PBS on MASK_S haplotypes, as we demonstrated MASK_S to be a close proxy for the South Asian component of contemporary South Asian genomes in REF I (see Materials and Methods). We detected selection signals for SNPs within 50 kb range of *HLA-G, HLA-F, TRIM31* and *TRIM40* genes linked to immune system (Ishitani et al. 2003; Rajagopalan and Long 2012; Fu et al. 2017; Liu et al. 2017; Zhao et al. 2017; Lin and Yan 2019), *KCTD6* linked to sweet taste signalling (Liu et al. 2013), *ACOX2* linked to branched fatty acid processing (Bjørklund et al. 2015; Vilarinho et al. 2016), *DOK5* linked to signal transduction (Favre et al. 2003) and *CDH4* possibly linked to brain segmentation (Babb et al. 2005).

However, especially *HLA* related signals must be taken with a grain of salt as genotyping for those regions is known to be problematic, which in turn might have partly affected the local ancestry deconvolution process. It is also essential to note that even though an initial ancestry deconvolution might improve detection compared to using admixed South Asian genomes, these selection signals might still be false positives.

5.3. Origins of the Eurasian admixture in Ethiopian genomes (REF III)

Similarly to South Asian genomes, Ethiopian genomes can be characterized by two major components; a local African (AF) and a Non-African (NAF) that arrived during post Neolithic migrations from West Eurasia. This study delves into possible sources for the NAF component using local ancestry inference-based approaches, akin to REF I and II, on modern Ethiopian populations.

5.3.1. Disassembling the Ethiopian genome

After creating masked non African (NAF) and African (AF) haplotypes from available data for contemporary Ethiopian genomes via PCAdmix, we first tested the assumption that Eurasian component is due to a single major admixture event that took place 3,000 years ago. We didn't find signatures for multiple waves except for Wolayta population which showed an additional more recent gene flow. Confirming the relatively homogenous nature of NAF component, we projected masked NAF onto PCA space obtained through whole genomes. PCA indicated Anatolian Neolithic (Anatolia_N), Minoans and North African Jews to be the closest proxies for NAF (**Figure 7**)**.** We further confirmed this affinity with outgroup f3 statistics. Although whole Ethiopian genomes showed higher affinity to Levant than Anatolia (as also reported by Lazaridis et al. 2016), the masked NAF component showed the opposite pattern. Additionally, based on top scoring populations for NAF ancestry in outgroup f3 analysis, Minoans were as close as Anatolia Neolithic samples (Z score <1) to the NAF ancestry. North African Jews, in particular Tunisian Jews, also showed high affinity to the Ethiopian NAF.

To further inspect the Eurasian component of Ethiopians in finer detail, we also performed f4 statistics as f4(Pop A, Pop B; Test, Mbuti) with Pop A and Pop B as candidates and Test as Amhara (representative whole Ethiopian genomes) or Amhara NAF (masked NAF haplotypes). Whole genome Amhara were significantly closer to Levantine ancestry than Anatolian one (closer to Levantine Chalcolithic than Levantine Neolithic) whereas the NAF component showed higher allele sharing with Anatolian Neolithic again. It is important to underline that these findings should not be interpreted as a direct link between Ethiopian NAF component and Anatolia Neolithic, but rather as a proximity between these two considering available modern and ancient genomes.

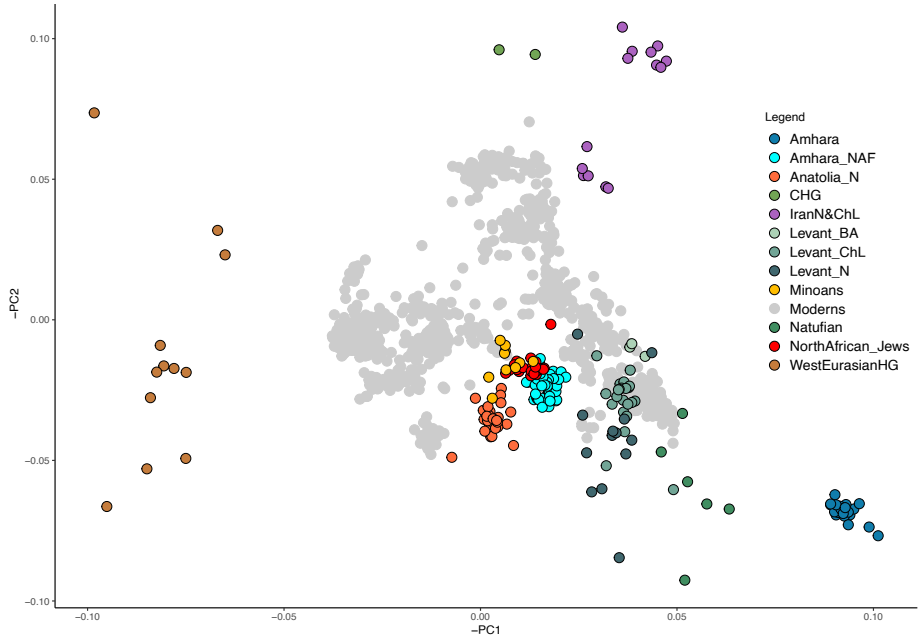

Figure 7. PCA of modern West Eurasian populations along with ancient genomes and masked NAF haplotypes. PCA was performed with modern genomes (grey) and ancient and masked NAF were projected onto this PCA space. Figure from Figure 2 (Molinaro et al. 2019), licenced under CC BY 4.0.

5.3.2. Possible link to Sea People

As Minoans and Tunisian Jews appeared to be the closest proxies for the Ethiopian NAF component, a tentative link between these three groups on a historical trade route bridging Crete (cradle of Minoan culture) with the Levant (Cross and Stager 2006; Ben-Shlomo et al. 2011; von Rüden 2014) and a seafarer nomad group in the Mediterrenean called the Sea People could be made. Although the origins of these nomads is still a topic of debate, it is known that they fought the Hittites, Egyptians and Levantines and were possibly settled in Palestine in the first millennium BCE (Bryce 2010). According to Egyptian inscriptions, the Sea People settled in Palestine were Denyen, which was supposedly linked to the tribe of Dan (from which Ethiopian Jews are thought to descend from) and Peleset tribes which were linked to the Philistines from the Levant (Cline 2014; D'Amato et al. 2015). These connections may point towards a contribution of the Sea People to an interim Minoan-like ancestry in the Levant, which arrived in Ethiopia ~3000 years ago through stepwise movements from that area. In addition, some Iron Age samples with high Anatolian-like ancestry have recently been discovered in the Levant, appearing and disappearing in the archeological record within a few century span (Feldman et al. 2019). These samples can be modelled as having 80% Anatolian Neolithic ancestry and ~20% Iranian Neolithic or CHG components and are similar to Ethiopian NAF with f4 statistics. This further strengthens the possible link between Ethiopian NAF and the Sea People.

6. CONCLUSIONS

The results presented in this thesis show that:

- For admixed populations with deeply divergent ancestries, such as South Asian and Ethiopian populations, local ancestry deconvolution methods can be effectively used to create proxies for the ancestral components which can be utilized for demographic and functional analyses (REF I, REF III).
- West Eurasian ancestry is likely to have arrived in South Asia in multiple waves and the South Asian component can be defined as a three-way split with East Asia and Andamanese ancestries (REF I).
- High frequency West Eurasian-specific functional alleles in *SLC45A2* and *SLC24A5* genes related to skin pigmentation in Europeans show post-admixture purifying and positive selective pressures in South Asian genomes, respectively (REF I).
- Before performing PBS, a preliminary local ancestry deconvolution step can improve selection detection and help with identifying the source of selection in admixed populations, especially by reducing false positive signals (REF II).
- XP-EHH signals detected in admixed populations can reflect selection pressures that acted on one of the ancestral populations instead of reflecting selection on the admixed population. Additionally, XP-EHH might produce pseudo selection signals due to shifts in allele frequencies related to admixture events. Therefore, XP-EHH must be used cautiously on admixed populations (REF II).
- The Eurasian component of Ethiopians has higher affinity to Anatolian Neolithic ancestry instead of previously suggested Levantine, with Minoans and Tunisian Jews being the closest available proxies (REF III).

7. EXTRAS

7.1. Creating artificial genomes for population genetics (REF X)

The sample data used in the research articles of this thesis were accessible from their corresponding publications. However, the vast majority of human genomic data are either not publicly available or subject to lengthy application procedures and ethical clearances related to valid privacy concerns. This creates a scientific barrier for researchers in terms of time and resources. To overcome the problem without infringing privacy of genetic donors, we tested the potential of generative neural networks for creating realistic artificial genomes (AGs) (Yelmen et al. 2021). We implemented a generative neural network (GAN) (Goodfellow et al. 2014) and a restricted Boltzmann machine (RBM) (Smolensky 1986, Teh et al. 2001) model for human SNP data. Due to computational and algorithmic limitations, we were able to generate short genome chunks instead of whole genomes. We demonstrated that these chunks (AGs) are capable of imitating real genomes in terms of allele frequencies, population structure, linkage blocks, pairwise haplotype distances and natural selection signals, with little to none privacy leakage from original training genomes. Additionally, we showed that AGs can be used for improving imputation quality of low frequency alleles when used as augmentations to real reference panels, and RBM latent space can be utilized to encode the real genomic data for further analysis of the training dataset. With further improvements to the models we proposed, AGs have the potential to be used as high quality, anonymous and easily accessible surrogates of genomic databases, hence eliminating a significant obstacle from scientific research.

8. APPENDIX

Table 3. Samples used in the three studies with their corresponding sources.

Population	Sample size	Source	Population	Sample size	Source
Birhor	16	Basu et al. 2016	Algerian	7	Lazaridis et al.2014
Brahmin_Guj	20	Basu et al. 2016	Amhara	24	Pagani et al. 2015
Irula	20	Basu et al. 2016	BantuKenya	6	Patterson et al. 2012
Iyer	19	Basu et al. 2016	BantuSA	8	Patterson et al. 2012
Jamatia	18	Basu et al. 2016	Biaka	20	Patterson et al. 2012
Jarawa	19	Basu et al. 2016	Canary_Islander	2	Lazaridis et al. 2014
Khatri	19	Basu et al. 2016	Egyptian	100	Pagani et al. 2015
Maratha	7	Basu et al. 2016	Egyptian	12	Behar et al. 2010
Onge	17	Basu et al. 2016	Esan	8	Lazaridis et al. 2014
Pallan	20	Basu et al. 2016	Esomali	24	Pagani et al. 2015
Paniya	18	Basu et al. 2016	Ethiopians	19	Behar et al. 2010
Santhal	20	Basu et al. 2016	Gambian	6	Lazaridis et al. 2014
UP_Low_Caste	5	Metspalu et al. 2011	Gumuz	24	Pagani et al. 2015
Chamar	10	Metspalu et al. 2011	Hadza	5	Pickrell et al. 2012
Dharkars	11	Metspalu et al. 2011	Jew_Ethiopian	7	Lazaridis et al. 2014
Dusadh	6	Metspalu et al. 2011	Jew_Ethiopian	13	Behar et al. 2010
Kanjars	8	Metspalu et al. 2011	Jew_Libyan	9	Lazaridis et al. 2014
Gujjar	15	Pathak et al. 2018	Jew_Moroccan	6	Lazaridis et al. 2014
Kamboj	14	Pathak et al. 2018	Jew_Moroccan	16	Behar et al. 2010
Ror	15	Pathak et al. 2018	Jew_Turkish	8	Lazaridis et al. 2014
Gujaratis	100	Hapmap3	Jew Tunisian	7	Lazaridis et al. 2014
Burusho	25	Li et al. 2008	Ju_hoan_North	5	Patterson et al. 2012
Brahui	25	Li et al. 2008	Kenya_400BP	1	Skoglund et al. 2017
Sindhi	24	Li et al. 2008	Khomani	11	Lazaridis et al. 2014
Kalash	23	Li et al. 2008	Libyan	5	Lazaridis et al. 2014
Pathan	22	Li et al. 2008	Luhya	8	Lazaridis et al. 2014
Balochi	24	Li et al. 2008	Luo	8	Lazaridis et al. 2014
Makrani	25	Li et al. 2008	LWK	80	The 1000 Genomes Project
French	28	Li et al. 2008	Malawi_Chencherere_5200BP	2	Skoglund et al. 2017
Sardinians	28	Li et al. 2008	Malawi_Fingira_2500BP	1	Skoglund et al. 2017
Adygei	17	Li et al. 2008	Malawi_Fingira_6100BP	2	Skoglund et al. 2017
Palestinians	46	Li et al. 2008	Malawi_Hora_8100BP	2	Skoglund et al. 2017
Han	44	Li et al. 2008	Malawi_Hora_9000BP	1	Skoglund et al. 2017
Japanese	28	Li et al. 2008	Mandenka	17	Patterson et al. 2012
Yorubas	21	Li et al. 2008	Masai	12	Lazaridis et al. 2014
Hungarians	19	Behar et al. 2010	Mbuti	10	Patterson et al. 2012
Lezgins	18	Behar et al. 2010	Mende	8	Lazaridis et al. 2014
Armenians	19	Behar et al. 2010	Moroccans	10	Lazaridis et al. 2016
Turks	19	Behar et al. 2010	Moroccans	10	Behar et al. 2010
Jordanians	20	Behar et al. 2010	Mota	1	GallegoLM et al. 2015
Iranians	19	Behar et al. 2010	Oromo	24	Pagani et al. 2015
Georgians	10	Behar et al. 2013	Somali	13	Lazaridis et al. 2014
Turkmens	20	Yunusbayev et al. 2012	South_Africa_2100-1200BP	3	Skoglund et al. 2017
Tajiks	15	Yunusbayev et al. 2012	Tanzania_Luxmanda_3100BP	2	Skoglund et al. 2017
Chechens	20	Yunusbayev et al. 2012	Tanzania_Pemba	2	Skoglund et al. 2017
Germans	13	Yunusbayev et al. 2015	Tanzania_Zanzibar	2	Skoglund et al. 2017

Population	Sample size	Source	Population	Sample size	Source
Tunisian	8	Lazaridis et al. 2014	Iraqi_Jews	11	Behar et al. 2010
Wolayta	24	Pagani et al. 2015	Italian_North	20	Patterson et al. 2012
YRI	80	The 1000 Genomes Project	Italian_South	2	Lazaridis et al. 2014
Abkhasian	9	Lazaridis et al. 2014	Italian_South	2	Lazaridis et al. 2016
Anatolia_BA	1	Lazaridis et al. 2017	Jew_Georgian	7	Lazaridis et al. 2014
Anatolia_ChL	1	Lazaridis et al. 2016	Jew_Georgian	4	Behar et al. 2010
Anatolia_EBA	2	Lazaridis et al. 2017	Jew_Iranian	9	Lazaridis et al. 2014
Anatolia_N	24	Mathieson et al. 2015	Jew_Iranian	4	Behar et al. 2010
Anatolia_N	9	Kilinc et al. 2016	Jew_Iraqi	6	Lazaridis et al. 2014
Anatolia_N	2	Omrak et al. 2016	Jew_Iraqi	11	Behar et al. 2010
Anatolia_N	2	Hofmanova et al. 2016	Jew_Yemenite	8	Lazaridis et al. 2014
Armenia_ChL	5	Lazaridis et al. 2016	Jew_Yemenite	15	Behar et al. 2010
Armenia_EBA	3	Lazaridis et al. 2016	Lebanese	7	Behar et al. 2010
Armenia_MLBA	8	Allentoft et al. 2015	Levant_BA	3	Lazaridis et al. 2016
Armenia_MLBA	1	Lazaridis et al. 2016	Levant_ChL	21	Harney et al. 2018
Armenian	10	Lazaridis et al. 2014	Levant_N	13	Lazaridis et al. 2016
Armenian	19	Behar et al. 2010	Maltese	8	Lazaridis et al. 2014
Bedouin	44	Patterson et al. 2012	Natufian	3	Lazaridis et al. 2016
CEU	90	The 1000 Genomes Project	Palestinian	38	Patterson et al. 2012
CHG	2	Jones et al. 2015	Sardinian	27	Patterson et al. 2012
Croatian	10	Lazaridis et al. 2014	Saudis	20	Behar et al. 2010
Cypriot	8	Lazaridis et al. 2014	SHG	6	Mathieson et al. 2015
Druze	39	Patterson et al. 2012	Sicilian	11	Lazaridis et al. 2014
EHG	3	Mathieson et al. 2015	Sidon_BA	5	Haber et al. 2017
English	10	Lazaridis et al. 2014	Spanish	53	Lazaridis et al. 2014
Europe_EN	1	Lazaridis et al. 2014	Spanish_North	4	Lazaridis et al. 2014
Europe_EN	21	Mathieson et al. 2015	Steppe_BA	30	Mathieson et al. 2015
Europe_EN	6	Lipson et al. 2017	Steppe_BA	18	Allentoft et al. 2015
Europe_MNChL	21	Mathieson et al. 2015	Steppe_Eneolithic	3	Mathieson et al. 2015
Europe_MNChL	3	Allentoft et al. 2015	Steppe_IA	1	Mathieson et al. 2015
French	29	Lazaridis et al. 2016	Syrians	16	Behar et al. 2010
French	25	Patterson et al. 2012	Ukrainian	9	Lazaridis et al. 2014
Georgians	20	Behar et al. 2010	Villabruna	1	Fu et al. 2016
Greece_Minoan_Lassithi	5	Lazaridis et al. 2017	WHG	1	Lazaridis et al. 2014
Greece_Minoan_Odigitria	5	Lazaridis et al. 2017	WHG	1	Mathieson et al. 2015
Greece_Mycenaean	4	Lazaridis et al. 2017	WHG	1	Mathieson et al. 2015
Greece_N	3	Hofmanova et al. 2016	Yemenese	10	Behar et al. 2010
Greece_N	1	Lazaridis et al. 2017	Darra_i_kur_MBA	1	Narasimhan et al. 2018
Greek	20	Lazaridis et al. 2014	Belt_Cave_Iran_Mesolithic_LC	1	Narasimhan et al. 2018
Hungarian	14	Lazaridis et al. 2014	Ganj_Dareh_N	3	Narasimhan et al. 2018
IBS	49	The 1000 Genomes Project	Ganj_Dareh_N_father.of. I1947_father.of.I1952	1	Narasimhan et al. 2018
Iran_ChL	5	Lazaridis et al. 2016	Ganj_Dareh_N_son.of.I1946_ brother.of.I1947	1	Narasimhan et al. 2018
Iran_N	5	Lazaridis et al. 2016	Hajji_Firuz	5	Narasimhan et al. 2018
Iran_N	3	Broushaki et al. 2016	Hajji_Firuz_BA	1	Narasimhan et al. 2018
Iranian	20	Behar et al. 2010	Hajji_Firuz_C	1	Narasimhan et al. 2018
Iranian	30	Lazaridis et al. 2016	Indus_diaspora	2	Narasimhan et al. 2018
Shahr_I_Sokhta_BA1	2	Narasimhan et al. 2018	Sintashta_MLBA_o2	2	Narasimhan et al. 2018
Tepe_Hissar_C	12	Narasimhan et al. 2019	Sintashta_MLBA_o2_ brother.of.I1057	2	Narasimhan et al. 2018

Population	Sample size	Source	Population	Sample size	Source
Tepe_Hissar_C_LC	1	Narasimhan et al. 2019	Sintashta_MLBA_o3	2	Narasimhan et al. 2018
Ak_Moustafa_MLBA1	1	Narasimhan et al. 2019	Steppe_EMBA	24	Narasimhan et al. 2018
Alpamsa_BA_Alakul	1	Narasimhan et al. 2019	Steppe_MLBA_East	15	Narasimhan et al. 2018
Dali_EBA	1	Narasimhan et al. 2019	Steppe_MLBA_West	38	Narasimhan et al. 2018
Kairan_MLBA_o	2	Narasimhan et al. 2019	West_Siberia_N	3	Narasimhan et al. 2018
Kairan_MLBA_o_LC	1	Narasimhan et al. 2019	Dashti_Kozy_BA	3	Narasimhan et al. 2018
Kanai_MBA	1	Narasimhan et al. 2019	Sarazm_EN	2	Narasimhan et al. 2018
Kyzlbulak_MBA2	1	Narasimhan et al. 2019	BMAC	17	Narasimhan et al. 2018
Maitan_MLBA_Alakul_o	2	Narasimhan et al. 2019	Geoksiur_EN	10	Narasimhan et al. 2018
Oy_Dzhaylau_MLBA_o2	2	Narasimhan et al. 2019	Geoksiur_EN.1d.rel.of.S8502	2	Narasimhan et al. 2018
Satan_MLBA_Alakul	1	Narasimhan et al. 2019	Gonur1_BA_LC	16	Narasimhan et al. 2018
Steppe_MLBA_East	15	Narasimhan et al. 2019	Gonur1_BA_o	2	Narasimhan et al. 2018
Steppe_MLBA_West	23	Narasimhan et al. 2019	Gonur1_BA_o2	2	Narasimhan et al. 2018
Talapty_MLBA	1	Narasimhan et al. 2019	Gonur1_BA_sibling.I1781_LC	1	Narasimhan et al. 2018
Taldysay_MLBA1	1	Narasimhan et al. 2019	Indus_diaspora	1	Narasimhan et al. 2018
Taldysay_MLBA2	1	Narasimhan et al. 2019	Parkhai_EBA	1	Narasimhan et al. 2018
Unknown_MLBA	1	Narasimhan et al. 2019	Parkhai_EBA_LC	1	Narasimhan et al. 2018
Zevakinskiy_LBA	6	Narasimhan et al. 2019	Parkhai_EN	4	Narasimhan et al. 2018
Zevakinskiy_MLBA	1	Narasimhan et al. 2019	Parkhai_LBA	1	Narasimhan et al. 2018
Aligrama_IA	3	Narasimhan et al. 2019	Parkhai_LBA_o	1	Narasimhan et al. 2018
Butkara_IA	3	Narasimhan et al. 2019	Parkhai_MBA	1	Narasimhan et al. 2018
Butkara_IA_mother.I6549	1	Narasimhan et al. 2019	Sumbar_LBA	1	Narasimhan et al. 2018
Loebanr_IA_father.I6292	1	Narasimhan et al. 2019	Tepe_Anau_EN	3	Narasimhan et al. 2018
Pakistan_IA_Aligrama_all	1	Narasimhan et al. 2019	BMAC	26	Narasimhan et al. 2018
Saidu_Sharif_IA	11	Narasimhan et al. 2019	Dzharkutan1_BA_LC	1	Narasimhan et al. 2018
Saidu_Sharif_IA_LC	1	Narasimhan et al. 2019	Dzharkutan2_BA	2	Narasimhan et al. 2018
Saidu_Sharif_IA_o	1	Narasimhan et al. 2019	Kashkarchi_BA	2	Narasimhan et al. 2018
SPGT	36	Narasimhan et al. 2019	Sappali_Tepe_BA_o	1	Narasimhan et al. 2018
Udegram_IA_1d.rel.I6900	1	Narasimhan et al. 2019			
Udegram_IA_father.or. son.I1799	1	Narasimhan et al. 2019			
Udegram_IA_LC	2	Narasimhan et al. 2019			
Udegram_IA_son.I13262	1	Narasimhan et al. 2019			
Afanasievo_1d.rel.I3950	1	Narasimhan et al. 2019			
Afanasievo_son.I3388_ son.I3950_brother.I3949	1	Narasimhan et al. 2019			
Afanasievo_son.I3388_ son.I3950_brother.I6714	1	Narasimhan et al. 2019			
Krasnoyarsk_MLBA_ father_or_son.I6718	1	Narasimhan et al. 2019			
Krasnoyarsk_MLBA_o	1	Narasimhan et al. 2019			
Petrovka	3	Narasimhan et al. 2019			
Preobrazhenka_MLBA	1	Narasimhan et al. 2019			
Sintashta_MLBA_ 1d.rel.I1086	1	Narasimhan et al. 2019			
Sintashta_MLBA 1st.degree.rel.I1055	1	Narasimhan et al. 2019			
Sintashta_MLBA_ 1st.degre e.rel.I1084	1	Narasimhan et al. 2019			
Sintashta_MLBA_ brother.of.I1053	1	Narasimhan et al. 2019			
Sintashta_MLBA_o1	4	Narasimhan et al. 2019			

SUMMARY IN ESTONIAN

Iidsete Euraasia mõjude iseloomustamine tänapäeva inimgenoomides

Tänapäeva inimpopulatsioonide genoomset koostist on kujundanud mineviku demograafilised sündmused nagu ränded, pudelikaelad, asutajaefektid ja erinevate gruppide segunemine. Nende protsesside uurimine ja meie liigi ajaloo selgitamine pakuvad endiselt väljakutseid, kuid on muutunud lihtsamaks tänu sekveneerimistehnoloogiate ja vana DNA (aDNA) meetodite arengule ning uutele andmeanalüüsi ja modelleerimise meetoditele. Ehkki aDNA on nüüdseks selles ettevõtmises hädavajalik tööriist, on selle kättesaadavus piiratud, kuna erinevates geograafilistes asukohtades on DNA pikaajaliseks säilimiseks erinevad tingimused, mis põhjustab paljudel juhtudel nii kvaliteedi- kui kvantiteediprobleeme. DNA lagunemine toimub nende tingimustega seotud keemiliste protsesside tõttu ja on valdavam vanemates proovides. Lisaks ei säili aDNA soojemas kliimas tavaliselt üldse. Seetõttu uuritakse tänapäeva inimeste genoome meie mineviku mõistmiseks endiselt laialdaselt, kuna neis on jäljed nendest iidsetest sündmustest.

"Anatoomiliselt moodsaks" nimetatud fenotüübiliste tunnuste evolutsioon ulatub Aafrika fossiilide põhjal vähemalt 300 000 aasta taha. Enamik geneetilisest varieeruvusest väljaspool Aafrikat pärineb hiljutisest "Aafrikast välja" sündmusest, mille kaks peamist võimalikku levikuteed Euraasiasse olid kas üle Siinai poolsaare Levandi piirkonda ("põhjapoolne tee") ja/või üle Punase mere Bab el Mandebi väina Araabiasse ("lõunapoolne tee") 50 000–100 000 aastat tagasi. Praeguse arheoloogilise ja geneetilise tõendusmaterjali põhjal ei ole nende teede osas endiselt täielikku konsensust. Pärast Aafrikast välja sündmust mõjutasid maailma inimgenoomide varieeruvust erinevad populatsioonide lahknemised koos evolutsiooniliste jõududega nagu geneetiline triiv ja looduslik valik. Teatud mandripiirkondade nagu Euroopa, Lõuna-Aasia ja Aafrika tänapäeva genofondi tekkimise defineeriv periood oli neoliitikumi-järgne inimpopulatsioonide ränne ja segunemine. Lähis-Ida ja Euraasia stepi populatsioonid segunesid kohapealsete populatsioonidega üle kogu Lääne- ja Lõuna-Euraasia ning nende rännete pikk laine ulatus isegi Ida- ja Lõuna-Aafrikani.

Doktoritöö peamine eesmärk on tänapäeva inimgenoome kasutades uurida süvitsi iidseid Euraasia komponente tänapäeva inimpopulatsioonides, keskendudes Lõuna-Aasia ja Etioopia populatsioonidele. Ehkki aDNA on selleks juba väärtuslikuks osutunud, võib ka tänapäeva inimgenoome vaadelda andmete allikana, mis koosneb iidsete kihtide kombinatsioonist, mida saab "välja kaevata". Sellele vastavalt on näidatud, et Lõuna-Aasia genoomid koosnevad Lõuna-Aasia (S) ja Lääne-Euraasia (N) komponentidest, mis on viimase 10 000 aasta jooksul segunenud. Etiooplasi on iseloomustatud kui segu mitmest eristatavast Aafrika põlvnemiskomponendist ja Euraasia komponendist, mis pärineb 3000 aasta tagusest Lähis-Idast. Käesolev doktoritöö hõlmab lähemat uurimust Lõuna-Aasia populatsioonide kahe peamise komponendi demograafiast (REF I), metodoloogilisi lähenemisi valiku tuvastamise parandamiseks segunenud populat-

sioonides ja võimalikke valiku signaale Lõuna-Aasia genoomides (REF I, II), ning sissevaadet etiooplaste Euraasia komponendi päritollu (REF III) lokaalse põlvnemise tuletamise meetodite abil.

Käesoleva doktoritöö peamised järeldused on järgmised:

- Tugevalt erinevate põlvnemisallikatega segunenud populatsioonide puhul nagu Lõuna-Aasia ja Etioopia populatsioonid saab lokaalse põlvnemise selgitamise meetodeid efektiivselt kasutada põlvnemiskomponentide lähenduste loomiseks, mida saab kasutada demograafilistes ja funktsionaalsetes analüüsides.
- Lääne-Euraasia komponent jõudis Lõuna-Aasiasse tõenäoliselt mitme lainena ning Lõuna-Aasia komponenti võib defineerida kui kolmest lahknemist Ida-Aasia ja Andamani komponentidega.
- Kõrge sagedusega Lääne-Euraasia-spetsiifilised funktsionaalsed alleelid *SLC45A2* ja *SLC24A5* geenides, mis on seotud naha pigmentatsiooniga eurooplastel, näitavad vastavalt segunemisjärgset puhastumist ja positiivse valiku survet Lõuna-Aasia genoomides.
- Enne populatsiooni haru statistiku (PBS) arvutamist võib esialgne lokaalse põlvnemise selgitamise samm parandada valiku tuvastamist ja aidata kindlaks teha valiku allikat segunenud populatsioonides, eriti valepositiivsete signaalide vähendamise teel.
- Segunenud populatsioonides tuvastatud populatsioonidevahelise pikenenud haplotüübi homosügootsuse (XP-EHH) signaalid võivad peegeldada valikusurveid, mis toimisid ühele eellaspopulatsioonile, mitte valikut segunenud populatsioonis. Lisaks võib XP-EHH anda pseudo-valikusignaale segunemisega seotud alleelisageduste muutuste tõttu. Seega tuleb XP-EHH meetodit segunenud populatsioonide puhul kasutada ettevaatusega.
- Etiooplaste Euraasia komponent on lähedasem Anatoolia neoliitilisele komponendile kui varem pakutud Levandi allikale, lähimad kättesaadavad vasted on minoalased ja Tuneesia juudid.

REFERENCES

1000 Genomes Project Consortium, Adam Auton, Lisa D Brooks, Richard M Durbin, Erik P Garrison, Hyun Min Kang, Jan O Korbel, et al. 2015. “A Global Reference for Human Genetic Variation.” *Nature* 526 (7571): 68–74. https://doi.org/10.1038/nature15393.

Alexander, David H., John Novembre, and Kenneth Lange. 2009. “Fast Model-Based Estimation of Ancestry in Unrelated Individuals.” *Genome Research* 19 (9): 1655–64. https://doi.org/10.1101/gr.094052.109.

Allentoft, Morten E., Martin Sikora, Karl Göran Sjögren, Simon Rasmussen, Morten Rasmussen, Jesper Stenderup, Peter B. Damgaard, et al. 2015. “Population Genomics of Bronze Age Eurasia.” *Nature* 522 (7555): 167–172. https://doi.org/10.1038/nature14507.

Altshuler, David M., Richard A. Gibbs, Leena Peltonen, David M. Altshuler, Richard A. Gibbs, Leena Peltonen, Emmanouil Dermitzakis, et al. 2010. “Integrating Common and Rare Genetic Variation in Diverse Human Populations.” *Nature* 467 (7311): 52–58. https://doi.org/10.1038/nature09298.

Armitage, Simon J., Sabah A. Jasim, Anthony E. Marks, Adrian G. Parker, Vitaly I. Usik, and Hans Peter Uerpmann. 2011. “The Southern Route ‘out of Africa’: Evidence for an Early Expansion of Modern Humans into Arabia.” *Science* 331 (6016): 453–456. https://doi.org/10.1126/science.1199113.

Babb, Sherry G., Shannon M. Kotradi, Bijal Shah, Christin Chiappini-Williamson, Lauren N. Bell, Glen Schmeiser, Elbert Chen, Qin Liu, and James A. Marrs. 2005. “Zebrafish R-Cadherin (Cdh4) Controls Visual System Development and Differentiation.” *Developmental Dynamics* 233 (3): 930–945. https://doi.org/10.1002/dvdy.20431.

Bae, Christopher J., Katerina Douka, and Michael D. Petraglia. 2017. “On the Origin of Modern Humans: Asian Perspectives.” *Science* 358 (6368). https://doi.org/10.1126/science.aai9067.

Basu, Analabha, Neeta Sarkar-Roy, and Partha P. Majumder. 2016. “Genomic Reconstruction of the History of Extant Populations of India Reveals Five Distinct Ancestral Components and a Complex Structure.” *Proceedings of the National Academy of Sciences* 113 (6): 1594–99. https://doi.org/10.1073/pnas.1513197113.

Behar, Doron M., Bayazit Yunusbayev, Mait Metspalu, Ene Metspalu, Saharon Rosset, Jüri Parik, Siiri Rootsi, et al. 2010. “The Genome-Wide Structure of the Jewish People.” *Nature* 466 (7303): 238–42. https://doi.org/10.1038/nature09103.

Behar, Doron M, Mait Metspalu, Yael Baran, Naama M Kopelman, Bayazit Yunusbayev, Ariella Gladstein, Shay Tzur, et al. 2013. “No Evidence from Genome-Wide Data of a Khazar Origin for the Ashkenazi Jews.” *Human Biology* 85 (6): 859–900. https://doi.org/10.3378/027.085.0604.

Ben-Shlomo, David, Eleni Nodarou, and Jeremy B. Rutter. 2011. “Transport Stirrup Jars from the Southern Levant: New Light on Commodity Exchange in the Eastern Mediterranean.” *American Journal of Archaeology* 115 (3): 329–353. https://doi.org/10.3764/aja.115.3.0329.

Bjørklund, Sunniva Stordal, Vessela N. Kristensen, Michael Seiler, Surendra Kumar, Grethe I.Grenaker Alnæs, Yao Ming, John Kerrigan, et al. 2015. “Expression of an Estrogen-Regulated Variant Transcript of the Peroxisomal Branched Chain Fatty Acid Oxidase ACOX2 in Breast Carcinomas.” *BMC Cancer* 15 (1): 1–13. https://doi.org/10.1186/s12885-015-1510-8.

Bocquet-Appel, Jean Pierre. 2011. “When the World’s Population Took off: The Springboard of the Neolithic Demographic Transition.” *Science* 333 (6042): 560–61. https://doi.org/10.1126/science.1208880.

Brisbin, Abra, Katarzyna Bryc, Jake Byrnes, Fouad Zakharia, Larsson Omberg, Jeremiah Degenhardt, Andrew Reynolds, Harry Ostrer, Jason G. Mezey, and Carlos D. Bustamante. 2012. “PCAdmix: Principal Components-Based Assignment of Ancestry Along Each Chromosome in Individuals with Admixed Ancestry from Two or More Populations.” *Human Biology* 84 (4): 343–64. https://doi.org/10.3378/027.084.0401.

Broushaki, Farnaz, Mark G. Thomas, Vivian Link, Saioa López, Lucy van Dorp, Karola Kirsanow, Zuzana Hofmanová, et al. 2016. “Early Neolithic Genomes from the Eastern Fertile Crescent.” *Science* 353 (6298): 499–503. https://doi.org/10.1126/science.aaf7943.

Bryce, Trevor. 2010. “The Kingdom of the Hittites”. *Oxford University Press.*

Cavalli-Sforza, L.L., Menozzi, P., and Piazza, A. 1994. “The history and geography of human genes” *Princeton University Press*.

Chaubey, Gyaneshwer, Mait Metspalu, Ying Choi, Reedik Mägi, Irene Gallego Romero, Pedro Soares, Mannis Van Oven, et al. 2011. “Population Genetic Structure in Indian Austroasiatic Speakers: The Role of Landscape Barriers and Sex-Specific Admixture.” *Molecular Biology and Evolution* 28 (2): 1013–24. https://doi.org/10.1093/molbev/msq288.

Cline, Eric H. 2014. “1177 B.C.: The Year Civilization Collapsed.” *Princeton University Press*. https://doi.org/10.1215/10474552-3697876.

Crawford, Nicholas G., Derek E. Kelly, Matthew E. B. Hansen, Marcia H. Beltrame, Shaohua Fan, Shanna L. Bowman, Ethan Jewett, et al. 2017. “Loci Associated with Skin Pigmentation Identified in African Populations.” *Science* 358 (6365). https://doi.org/10.1126/science.aan8433.

Cross, Frank Moore, and Lawrence E. Stager. 2006. “Cypro-Minoan Inscriptions Found in Ashkelon.” *Israel Exploration Journal*.

D’Amato, R. & Salimbeti, A. 2015. “Sea Peoples of the Bronze Age Mediterranean c.1400 BC–1000 BC”. *Osprey Publishing*.

Derricourt, Robin. 2005. “Getting ‘Out of Africa’: Sea Crossings, Land Crossings and Culture in the Hominin Migrations.” *Journal of World Prehistory* 19 (2): 119–132. https://doi.org/10.1007/s10963-006-9002-z.

Diaz-Papkovich, Alex, Luke Anderson-Trocmé, Chief Ben-Eghan, and Simon Gravel. 2019. “UMAP Reveals Cryptic Population Structure and Phenotype Heterogeneity in Large Genomic Cohorts.” *PLoS Genetics* 15 (11): e1008432. https://doi.org/10.1371/journal.pgen.1008432.

Fairfax, Benjamin P., Seiko Makino, Jayachandran Radhakrishnan, Katharine Plant, Stephen Leslie, Alexander Dilthey, Peter Ellis, Cordelia Langford, Fredrik O. Vannberg, and Julian C. Knight. 2012. “Genetics of Gene Expression in Primary Immune Cells Identifies Cell Type-Specific Master Regulators and Roles of HLA Alleles.” *Nature Genetics* 44 (5): 502–10. https://doi.org/10.1038/ng.2205.

Favre, C., A. Gérard, E. Clauzier, P. Pontarotti, D. Olive, and Jacques A. Nunès. 2003. “DOK4 and DOK5: New Dok-Related Genes Expressed in Human T Cells.” *Genes and Immunity* 4 (1): 40–45. https://doi.org/10.1038/sj.gene.6363891.

Feldman, Michal, Daniel M. Master, Raffaela A. Bianco, Marta Burri, Philipp W. Stockhammer, Alissa Mittnik, Adam J. Aja, Choongwon Jeong, and Johannes Krause. 2019. “Ancient DNA Sheds Light on the Genetic Origins of Early Iron Age Philistines.” *Science Advances* 5 (7): eaax0061. https://doi.org/10.1126/sciadv.aax0061.

Figueiredo, Jane C., Li Hsu, Carolyn M. Hutter, Yi Lin, Peter T. Campbell, John A. Baron, Sonja I. Berndt, et al. 2014. "Genome-Wide Diet-Gene Interaction Analyses for Risk of Colorectal Cancer." *PLoS Genetics* 10 (4): e1004228. https://doi.org/10.1371/journal.pgen.1004228.

Fu, Binqing, Yonggang Zhou, Xiang Ni, Xianhong Tong, Xiuxiu Xu, Zhongjun Dong, Rui Sun, Zhigang Tian, and Haiming Wei. 2017. "Natural Killer Cells Promote Fetal Development through the Secretion of Growth-Promoting Factors." *Immunity* 47 (6): 1100–1113. https://doi.org/10.1016/j.immuni.2017.11.018.

Fu, Qiaomei, Cosimo Posth, Mateja Hajdinjak, Martin Petr, Swapan Mallick, Daniel Fernandes, Anja Furtwängler, et al. 2016. "The Genetic History of Ice Age Europe." *Nature* 534 (7606): 200–205. https://doi.org/10.1038/nature17993.

Gallego Llorente, M., E. R. Jones, A. Eriksson, V. Siska, K. W. Arthur, J. W. Arthur, M. C. Curtis, et al. 2015. "Ancient Ethiopian Genome Reveals Extensive Eurasian Admixture throughout the African Continent." *Science* 350 (6262): 820–822. https://doi.org/10.1126/science.aad2879.

Gelabert, Pere, Iñigo Olalde, Toni De-DIos, Sergi Civit, and Carles Lalueza-Fox. 2017. "Malaria Was a Weak Selective Force in Ancient Europeans." *Scientific Reports* 7 (1): 1–10. https://doi.org/10.1038/s41598-017-01534-5.

Goodfellow, Ij, Jean Pouget-Abadie, and Mehdi Mirza. 2014. "Generative Adversarial Networks (GANs) – Tutorial." *(NIPS) Neural Information Processing Systems*.

Guan, Yongtao. 2014. "Detecting Structure of Haplotypes and Local Ancestry." *Genetics* 196 (3): 625–642. https://doi.org/10.1534/genetics.113.160697.

Haak, Wolfgang, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick, Bastien Llamas, Guido Brandt, et al. 2015. "Massive Migration from the Steppe Was a Source for Indo-European Languages in Europe." *Nature* 522 (7555): 207–11. https://doi.org/10.1038/nature14317.

Haber, Marc, Claude Doumet-Serhal, Christiana Scheib, Yali Xue, Petr Danecek, Massimo Mezzavilla, Sonia Youhanna, et al. 2017. "Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences." *American Journal of Human Genetics* 101 (2): 274–282. https://doi.org/10.1016/j.ajhg.2017.06.013.

Harney, Éadaoin, Hila May, Dina Shalem, Nadin Rohland, Swapan Mallick, Iosif Lazaridis, Rachel Sarig, et al. 2018. "Ancient DNA from Chalcolithic Israel Reveals the Role of Population Mixture in Cultural Transformation." *Nature Communications* 9 (1): 1–11. https://doi.org/10.1038/s41467-018-05649-9.

Hodgson, Jason A., Connie J. Mulligan, Ali Al-Meeri, and Ryan L. Raaum. 2014. "Early Back-to-Africa Migration into the Horn of Africa." *PLoS Genetics* 10 (6): e1004393. https://doi.org/10.1371/journal.pgen.1004393.

Hodgson, Jason A., Joseph K. Pickrell, Laurel N. Pearson, Ellen E. Quillen, António Prista, Jorge Rocha, Himla Soodyall, Mark D. Shriver, and George H. Perry. 2014. "Natural Selection for the Duffy-Null Allele in the Recently Admixed People of Madagascar." *Proceedings of the Royal Society B: Biological Sciences* 281 (1789): 20140930. https://doi.org/10.1098/rspb.2014.0930.

Hofmanová, Zuzana, Susanne Kreutzer, Garrett Hellenthal, Christian Sell, Yoan Diekmann, David Díez-Del-Molino, Lucy Van Dorp, et al. 2016. "Early Farmers from across Europe Directly Descended from Neolithic Aegeans." *Proceedings of the National Academy of Sciences of the United States of America* 113 (25): 6886–6891. https://doi.org/10.1073/pnas.1523951113.

Hublin, Jean Jacques, Abdelouahed Ben-Ncer, Shara E. Bailey, Sarah E. Freidline, Simon Neubauer, Matthew M. Skinner, Inga Bergmann, et al. 2017. "New Fossils from Jebel Irhoud, Morocco and the Pan-African Origin of Homo Sapiens." *Nature* 546 (7657): 289–292. https://doi.org/10.1038/nature22336.

Hudson, Richard R. 2002. "Generating Samples under a Wright-Fisher Neutral Model of Genetic Variation." *Bioinformatics* 18 (2): 337–38. https://doi.org/10.1093/bioinformatics/18.2.337.

Huerta-Sánchez, Emilia, Michael DeGiorgio, Luca Pagani, Ayele Tarekegn, Rosemary Ekong, Tiago Antao, Alexia Cardona, et al. 2013. "Genetic Signatures Reveal High-Altitude Adaptation in a Set of Ethiopian Populations." *Molecular Biology and Evolution* 30 (8): 1877–1888. https://doi.org/10.1093/molbev/mst089.

Huerta-Sánchez, Emilia, Xin Jin, Asan, Zhuoma Bianba, Benjamin M. Peter, Nicolas Vinckenbosch, Yu Liang, et al. 2014. "Altitude Adaptation in Tibetans Caused by Introgression of Denisovan-like DNA." *Nature* 512 (7513): 194–197. https://doi.org/10.1038/nature13408.

Ishitani, Akiko, Noriko Sageshima, Ni Lee, Natalia Dorofeeva, Katsuhiko Hatake, Hans Marquardt, and Daniel E. Geraghty. 2003. "Protein Expression and Peptide Binding Suggest Unique and Interacting Functional Roles for HLA-E, F, and G in Maternal-Placental Immune Recognition." *The Journal of Immunology* 171 (3): 1376–1384. https://doi.org/10.4049/jimmunol.171.3.1376.

Jones, Eppie R., Gloria Gonzalez-Fortes, Sarah Connell, Veronika Siska, Anders Eriksson, Rui Martiniano, Russell L. McLaughlin, et al. 2015. "Upper Palaeolithic Genomes Reveal Deep Roots of Modern Eurasians." *Nature Communications* 6 (1): 1–8. https://doi.org/10.1038/ncomms9912.

Jouganous, Julien, Will Long, Aaron P. Ragsdale, and Simon Gravel. 2017. "Inferring the Joint Demographic History of Multiple Populations: Beyond the Diffusion Approximation." *Genetics* 206 (3): 1549–67. https://doi.org/10.1534/genetics.117.200493.

Kitchen, Andrew, Christopher Ehret, Shiferaw Assefa, and Connie J. Mulligan. 2009. "Bayesian Phylogenetic Analysis of Semitic Languages Identifies an Early Bronze Age Origin of Semitic in the Near East." *Proceedings of the Royal Society B: Biological Sciences* 276 (1668): 2703–2710. https://doi.org/10.1098/rspb.2009.0408.

Kılınç, Gülşah Merve, Ayça Omrak, Füsun Özer, Torsten Günther, Ali Metin Büyük-karakaya, Erhan Bıçakçı, Douglas Baird, et al. 2016. "The Demographic Development of the First Farmers in Anatolia." *Current Biology* 26 (19): 2659–2666. https://doi.org/10.1016/j.cub.2016.07.057.

Lazaridis, Iosif, Alissa Mittnik, Nick Patterson, Swapan Mallick, Nadin Rohland, Saskia Pfrengle, Anja Furtwängler, et al. 2017. "Genetic Origins of the Minoans and Mycenaeans." *Nature* 548 (7666): 214–218. https://doi.org/10.1038/nature23310.

Lazaridis, Iosif, Dani Nadel, Gary Rollefson, Deborah C. Merrett, Nadin Rohland, Swapan Mallick, Daniel Fernandes, et al. 2016. "Genomic Insights into the Origin of Farming in the Ancient Near East." *Nature* 536 (7617): 419–424. https://doi.org/10.1038/nature19310.

Lazaridis, Iosif, Nick Patterson, Alissa Mittnik, Gabriel Renaud, Swapan Mallick, Karola Kirsanow, Peter H. Sudmant, et al. 2014. "Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans." *Nature* 513 (7518): 409–413. https://doi.org/10.1038/nature13673.

Li, Jun Z., Devin M. Absher, Hua Tang, Audrey M. Southwick, Amanda M. Casto, Sohini Ramachandran, Howard M. Cann, et al. 2008. "Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation." *Science* 319 (5866): 1100–1104. https://doi.org/10.1126/science.1153717.

Lin, Aifen, and Wei Hua Yan. 2019. "The Emerging Roles of Human Leukocyte Antigen-F in Immune Modulation and Viral Infection." *Frontiers in Immunology* 10: 964. https://doi.org/10.3389/fimmu.2019.00964.

Lipson, Mark. 2020. "Applying F4-Statistics and Admixture Graphs: Theory and Examples." *Molecular Ecology Resources* 20 (6): 1658–1667. https://doi.org/10.1111/1755-0998.13230.

Lipson, Mark, Anna Szécsényi-Nagy, Swapan Mallick, Annamária Pósa, Balázs Stégmár, Victoria Keerl, Nadin Rohland, et al. 2017. "Parallel Palaeogenomic Transects Reveal Complex Genetic History of Early European Farmers." *Nature* 551 (7680): 368–372. https://doi.org/10.1038/nature24476.

Liu, Bingyu, Meng Zhang, Honglei Chu, Honghai Zhang, Haifeng Wu, Guanhua Song, Peng Wang, et al. 2017. "The Ubiquitin E3 Ligase TRIM31 Promotes Aggregation and Activation of the Signaling Adaptor MAVS through Lys63-Linked Polyubiquitination." *Nature Immunology* 18 (2): 214–224. https://doi.org/10.1038/ni.3641.

Liu, Zhepeng, Yaqian Xiang, and Guihong Sun. 2013. "The KCTD Family of Proteins: Structure, Function, Disease Relevance." *Cell and Bioscience* 3 (1): 1–5. https://doi.org/10.1186/2045-3701-3-45.

Loosdrecht, Marieke van de, Abdeljalil Bouzouggar, Louise Humphrey, Cosimo Posth, Nick Barton, Ayinuer Aximu-Petri, Birgit Nickel, et al. 2018. "Pleistocene North African Genomes Link Near Eastern and Sub-Saharan African Human Populations." *Science* 360 (6388): 548–552. https://doi.org/10.1126/science.aar8380.

Maples, Brian K., Simon Gravel, Eimear E. Kenny, and Carlos D. Bustamante. 2013. "RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference." *American Journal of Human Genetics* 93 (2): 278–88. https://doi.org/10.1016/j.ajhg.2013.06.020.

Martin, Michael D., Flora Jay, Sergi Castellano, and Montgomery Slatkin. 2017. "Determination of Genetic Relatedness from Low-Coverage Human Genome Sequences Using Pedigree Simulations." *Molecular Ecology* 26 (16): 4145–4157. https://doi.org/10.1111/mec.14188.

Mathieson, Iain, Iosif Lazaridis, Nadin Rohland, Swapan Mallick, Nick Patterson, Songül Alpaslan Roodenberg, Eadaoin Harney, et al. 2015. "Genome-Wide Patterns of Selection in 230 Ancient Eurasians." *Nature* 528 (7583): 499–503. https://doi.org/10.1038/nature16152.

McInnes, Leland, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. "UMAP: Uniform Manifold Approximation and Projection." *Journal of Open Source Software*. https://doi.org/10.21105/joss.00861.

Mellars, P., K. C. Gori, M. Carr, P. A. Soares, and M. B. Richards. 2013. "Genetic and Archaeological Perspectives on the Initial Modern Human Colonization of Southern Asia." *Proceedings of the National Academy of Sciences* 110 (26): 10699–704. https://doi.org/10.1073/pnas.1306043110.

Metspalu, Mait, Irene Gallego Romero, Bayazit Yunusbayev, Gyaneshwer Chaubey, Chandana Basu Mallick, Georgi Hudjashov, Mari Nelis, et al. 2011. "Shared and Unique Components of Human Population Structure and Genome-Wide Signals of Positive Selection in South Asia." *American Journal of Human Genetics* 89 (6): 731–44. https://doi.org/10.1016/j.ajhg.2011.11.010.

Molinaro, Ludovica, Francesco Montinaro, Burak Yelmen, Davide Marnetto, Doron M. Behar, Toomas Kivisild, and Luca Pagani. 2019. "West Asian Sources of the Eurasian Component in Ethiopians: A Reassessment." *Scientific Reports* 9 (1): 1–8. https://doi.org/10.1038/s41598-019-55344-y.

Mörseburg, Alexander, Luca Pagani, Francois-xavier Ricaut, Bryndis Yngvadottir, Eadaoin Harney, Cristina Castillo, Tom Hoogervorst, et al. 2016. "Multi-Layered Population Structure in Island Southeast Asians Multi-Layered Population Structure in Island Southeast Asians." *European Journal of Human Genetics* 24 (11): 1605–1611. https://doi.org/10.1038/ejhg.2016.60.

Narasimhan, Vagheesh M., Nick Patterson, Priya Moorjani, Nadin Rohland, Rebecca Bernardos, Swapan Mallick, Iosif Lazaridis, et al. 2019. "The Formation of Human Populations in South and Central Asia." *Science* 365 (6457). https://doi.org/10.1126/science.aat7487.

Omrak, Ayça, Torsten Günther, Cristina Valdiosera, Emma M. Svensson, Helena Malmström, Henrike Kiesewetter, William Aylward, Jan Storå, Mattias Jakobsson, and Anders Götherström. 2016. "Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool." *Current Biology* 26 (2): 270–275. https://doi.org/10.1016/j.cub.2015.12.019.

Pagani, Luca, Toomas Kivisild, Ayele Tarekegn, Rosemary Ekong, Chris Plaster, Irene Gallego Romero, Qasim Ayub, et al. 2012. "Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool." *American Journal of Human Genetics* 91 (1): 83–96. https://doi.org/10.1016/j.ajhg.2012.05.015.

Pagani, Luca, Stephan Schiffels, Deepti Gurdasani, Petr Danecek, Aylwyn Scally, Yuan Chen, Yali Xue, et al. 2015. "Tracing the Route of Modern Humans out of Africa by Using 225 Human Genome Sequences from Ethiopians and Egyptians." *American Journal of Human Genetics* 96 (6): 986–991. https://doi.org/10.1016/j.ajhg.2015.04.019.

Palmer, Nicholette D., Mark O. Goodarzi, Carl D. Langefeld, Nan Wang, Xiuqing Guo, Kent D. Taylor, Tasha E. Fingerlin, et al. 2015. "Genetic Variants Associated with Quantitative Glucose Homeostasis Traits Translate to Type 2 Diabetes in Mexican Americans: The GUARDIAN (Genetics Underlying Diabetes in Hispanics) Consortium." *Diabetes* 64 (5): 1853–66. https://doi.org/10.2337/db14-0732.

Pathak, Ajai K, Anurag Kadian, Alena Kushniarevich, Francesco Montinaro, Mayukh Mondal, Linda Ongaro, Manvendra Singh, et al. 2018. "The Genetic Ancestry of Modern Indus Valley Populations from Northwest India." *American Journal of Human Genetics* 103 (6): 918–29. https://doi.org/10.1016/j.ajhg.2018.10.022.

Patterson, Nick, Priya Moorjani, Yontao Luo, Swapan Mallick, Nadin Rohland, Yiping Zhan, Teri Genschoreck, Teresa Webster, and David Reich. 2012. "Ancient Admixture in Human History." *Genetics* 192 (3): 1065–93. https://doi.org/10.1534/genetics.112.145037.

Pickrell, Joseph K., Nick Patterson, Chiara Barbieri, Falko Berthold, Linda Gerlach, Tom Güldemann, Blesswell Kure, et al. 2012. "The Genetic Prehistory of Southern Africa." *Nature Communications* 3 (1): 1–6. https://doi.org/10.1038/ncomms2140.

Pickrell, Joseph K., Nick Patterson, Po Ru Loh, Mark Lipson, Bonnie Berger, Mark Stoneking, Brigitte Pakendorf, and David Reich. 2014. "Ancient West Eurasian Ancestry in Southern and Eastern Africa." *Proceedings of the National Academy of Sciences of the United States of America* 111 (7): 2632–2637. https://doi.org/10.1073/pnas.1313787111.

Pritchard, Jonathan K., Matthew Stephens, and Peter Donnelly. 2000. “Inference of Population Structure Using Multilocus Genotype Data.” *Genetics* 155 (2): 945–959. https://doi.org/10.1093/genetics/155.2.945.

Raghavan, Maanasa, Pontus Skoglund, Kelly E. Graf, Mait Metspalu, Anders Albrechtsen, Ida Moltke, Simon Rasmussen, et al. 2014. “Upper Palaeolithic Siberian Genome Reveals Dual Ancestry of Native Americans.” *Nature* 505 (7481): 87–91. https://doi.org/10.1038/nature12736.

Rajagopalan, Sumati, and Eric O. Long. 2012. “Cellular Senescence Induced by CD158d Reprograms Natural Killer Cells to Promote Vascular Remodeling.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (50): 20596–20601. https://doi.org/10.1073/pnas.1208248109.

Reich, David, Kumarasamy Thangaraj, Nick Patterson, Alkes L. Price, and Lalji Singh. 2009. “Reconstructing Indian Population History.” *Nature* 461 (7263): 489–94. https://doi.org/10.1038/nature08365.

Reyes-Centeno, Hugo, Mark Hubbe, Tsunehiko Hanihara, Chris Stringer, and Katerina Harvati. 2015. “Testing Modern Human Out-of-Africa Dispersal Models and Implications for Modern Human Origins.” *Journal of Human Evolution* 87: 95–106. https://doi.org/10.1016/j.jhevol.2015.06.008.

Rüden, Constance von. 2014. “Beyond the East-West Dichotomy in Syrian and Levantine Wall Paintings.” In *Critical Approaches to Ancient Near Eastern Art* 10 (9781614510352.55). https://doi.org/10.1515/9781614510352.55.

Sabeti, Pardis C., Patrick Varilly, Ben Fry, Jason Lohmueller, Elizabeth Hostetter, Chris Cotsapas, Xiaohui Xie, et al. 2007. “Genome-Wide Detection and Characterization of Positive Selection in Human Populations.” *Nature* 449 (7164): 913–918. https://doi.org/10.1038/nature06250.

Salter-Townshend, Michael, and Simon Myers. 2019. “Fine-Scale Inference of Ancestry Segments without Prior Knowledge of Admixing Groups.” *Genetics* 212 (3): 869–889. https://doi.org/10.1534/genetics.119.302139.

Sheehan, Sara, and Yun S. Song. 2016. “Deep Learning for Population Genetic Inference.” *PLoS Computational Biology* 12 (3): e1004845. https://doi.org/10.1371/journal.pcbi.1004845.

Skoglund, Pontus, Swapan Mallick, Maria Cátira Bortolini, Niru Chennagiri, Tábita Hünemeier, Maria Luiza Petzl-Erler, Francisco Mauro Salzano, Nick Patterson, and David Reich. 2015. “Genetic Evidence for Two Founding Populations of the Americas.” *Nature* 525 (7567): 104–108. https://doi.org/10.1038/nature14895.

Skoglund, Pontus, Jessica C. Thompson, Mary E. Prendergast, Alissa Mittnik, Kendra Sirak, Mateja Hajdinjak, Tasneem Salie, et al. 2017. “Reconstructing Prehistoric African Population Structure.” *Cell* 171 (1): 59–71. https://doi.org/10.1016/j.cell.2017.08.049.

Smolensky, P. 1986. “Information Processing in Dynamical Systems: Foundations of Harmony Theory.” *Parallel Distributed Processing Explorations in the Microstructure of Cognition*.

Teh, Yee Whye, and Geoffrey E. Hinton. 2001. “Rate-Coded Restricted Boltzmann Machines for Face Recognition.” *Advances in Neural Information Processing Systems* 908–914.

Vilarinho, Sílvia, Sinan Sari, Francesca Mazzacuva, Kaya Bilgüvar, Güldal Esendagli-Yilmaz, Dhanpat Jain, Gülen Akyol, et al. 2016. “ACOX2 Deficiency: A Disorder of Bile Acid Synthesis with Transaminase Elevation, Liver Fibrosis, Ataxia, and Cognitive Impairment.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (40): 11289–11293. https://doi.org/10.1073/pnas.1613228113.

Vitti, Joseph J., Sharon R. Grossman, and Pardis C. Sabeti. 2013. "Detecting Natural Selection in Genomic Data." *Annual Review of Genetics* 47: 97–120. https://doi.org/10.1146/annurev-genet-111212-133526.

Wagh, Kshitij, Aatish Bhatia, Gabriela Alexe, Anupama Reddy, Vijay Ravikumar, Michael Seiler, Michael Boemo, et al. 2012. "Lactase Persistence and Lipid Pathway Selection in the Maasai." *PLoS ONE* 7 (9): e44751. https://doi.org/10.1371/journal.pone.0044751.

Walford, Geoffrey A., Stefan Gustafsson, Denis Rybin, Alena Stancáková, Han Chen, Ching Ti Liu, Jaeyoung Hong, et al. 2016. "Genome-Wide Association Study of the Modified Stumvoll Insulin Sensitivity Index Identifies BCL2 and FAM19A2 as Novel Insulin Sensitivity Loci." *Diabetes* 65 (10): 3200–3211. https://doi.org/10.2337/db16-0199.

Williams, Amy. 2016. "Admix-Simu: Program to Simulate Admixture between Multiple Populations." *Zenodo*. https://doi.org/10.5281/zenodo.45517. Accessed November 1, 2020.

Yelmen, Burak, Aurelien Decelle, Linda Ongaro, Davide Marnetto, Corentin Tallec, Francesco Montinaro, Cyril Furtlehner, Luca Pagani, and Flora Jay. 2021. "Creating Artificial Human Genomes Using Generative Neural Networks." *PLoS Genetics* 17 (2): e1009303. https://doi.org/10.1371/journal.pgen.1009303.

Yelmen, Burak, Davide Marnetto, Ludovica Molinaro, Rodrigo Flores, Mayukh Mondal, and Luca Pagani. 2021. "Improving Selection Detection with Population Branch Statistic on Admixed Populations." *Genome Biology and Evolution* 13 (4): evab039. https://doi.org/10.1093/gbe/evab039.

Yelmen, Burak, Mayukh Mondal, Davide Marnetto, Ajai K. Pathak, Francesco Montinaro, Irene Gallego Romero, Toomas Kivisild, Mait Metspalu, Luca Pagani, and Evelyne Heyer. 2019. "Ancestry-Specific Analyses Reveal Differential Demographic Histories and Opposite Selective Pressures in Modern South Asian Populations." *Molecular Biology and Evolution* 36 (8): 1628–1642. https://doi.org/10.1093/molbev/msz037.

Yi, Xin, Yu Liang, Emilia Huerta-Sanchez, Xin Jin, Zha Xi Ping Cuo, John E. Pool, N. Xu, et al. 2010. "Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude." *Science* 329 (5987): 75–78. https://doi.org/10.1126/science.1190371.

Yunusbayev, Bayazit, Mait Metspalu, Mari Järve, Ildus Kutuev, Siiri Rootsi, Ene Metspalu, Doron M. Behar, et al. 2012. "The Caucasus as an Asymmetric Semipermeable Barrier to Ancient Human Migrations." *Molecular Biology and Evolution* 29 (1): 359–65. https://doi.org/10.1093/molbev/msr221.

Yunusbayev, Bayazit, Mait Metspalu, Ene Metspalu, Albert Valeev, Sergei Litvinov, Ruslan Valiev, Vita Akhmetova, et al. 2015. "The Genetic Legacy of the Expansion of Turkic-Speaking Nomads across Eurasia." *PLoS Genetics* 11 (4): 1–24. https://doi.org/10.1371/journal.pgen.1005068.

Zhao, Chunyuan, Mutian Jia, Hui Song, Zhongxia Yu, Wenwen Wang, Qi Li, Lining Zhang, Wei Zhao, and Xuetao Cao. 2017. "The E3 Ubiquitin Ligase TRIM40 Attenuates Antiviral Immune Responses by Targeting MDA5 and RIG-I." *Cell Reports* 21 (6): 1613–1623. https://doi.org/10.1016/j.celrep.2017.10.020.

ACKNOWLEDGEMENTS

I'd like to start by thanking my supervisors Luca Pagani, Mait Metspalu and Toomas Kivisild for all the guidance and wisdom they provided. This thesis wouldn't become a reality without them. Luca, as my primary supervisor, has been a great support for all my research initiatives and very understanding of my untimely botherings. Mait, likewise, helped me greatly in overcoming both scientific and bureaucratic obstacles. Toomas provided invaluable insights especially during the first part of my studies and has been a great source of information and feedback. I'm also very thankful to all the co-authors in our studies for all the work and help, it wouldn't be possible without them. Also, I want to thank all the people in EBC for creating a really nice work environment full of good ideas. I want to especially thank Mayukh Mondal, Davide Marnetto, Linda Ongaro, Francesco Montinaro, Ajai Kumar Pathak and Ludovica Molinaro for their professional input and friendship. I'm also very thankful to Flora Jay for all the help and shaping the direction of my future studies.

I want to thank my family for their support, especially my mother who has been the most important person in my life. I'm also grateful to Prof. Dr. Hüseyin Demirörs for the invaluable help with my mother's health. Last but not least, I want to express my deepest gratitudes to my friends, especially Mehmet Tarık Demir, Dr. Enver Ahmet Demir, Kaan Özhan, Katre Salu and Joonas Karl Kuusik for their fellowship and being there for me in my hard times.

Thanks to High Performance Computing Center of the University of Tartu for providing computational resources. Thanks to European Regional Development Fund, ASTRA and Mobilitas Pluss projects for the funding. Additionally, I want to thank Archimedes Foundation for the Dora Plus mobility scholarship. I'm also very thankful to Estonian Biocentre, Institute of Genomics for the financial assistance and Institute of Molecular and Cell Biology for the help during my studies.

PUBLICATIONS

CURRICULUM VITAE

Date of Birth: March 18, 1990
Place of Birth: İzmir, Turkey
Citizenship: Turkish

Education

University of Tartu, Institute of Molecular and Cell Biology, Gene Technology, **Ph.D.** (2017–…)

Middle East Technical University, Department of Biological Sciences, **M.Sc.** (2014–2017)

Middle East Technical University, Department of Molecular Biology and Genetics, **B.Sc.** (2008–2014)

Ankara Atatürk Anadolu High School (2004–2008)

Professional experience

University of Tartu, Institute of Genomics, **Junior research fellow of modern population genetics** (2017–…)

IQ Smart Software Solutions, **Software developer** (2016)

TUBITAK (The Scientific and Technological Research Council of Turkey) Research Projects, Molecular Phylogenetics of Willow Species and Characterization of Gene Resources of Economically Important Species (*Salix alba* and *Salix excelsa*) for Improvement, **Scholarship student** (2014–2015)

INTERGEN Genetics and Rare Diseases Diagnosis, Research & Application Center – **Intern** (2013)

Teaching experience

University of Tartu, Chair of Evolutionary Biology, Evolution and the Natural World – (2020/2021 Autumn)

University of Tartu, Chair of Evolutionary Biology, Evolution and the Natural World – (2019/2020 Autumn)

University of Tartu, Chair of Evolutionary Biology, Evolutionary Processes – (2018/2019 Spring)

University of Tartu, Chair of Evolutionary Biology, Evolution and the Natural World – (2018/2019 Autumn)

University of Tartu, Chair of Evolutionary Biology, Evolutionary Processes – (2017/2018 Spring)

Mentoring

Gleb Kovalev (B.Sc.) – **Thesis supervisor** (2020–2021)

Annabella Kadavanich (B.Sc.) – **Intern supervisor** (2020)

Languages

Turkish (Native)

English (Fluent)

Conference talks & posters

ESHG European Human Genetics Virtual Conference – Interactive e-poster: "Creating Artificial Human Genomes Using Generative Neural Networks" (2020)

IMCB and IG Annual Conference, Tartu, Estonia – Talk: "Creating Artificial Human Genomes Using Generative Neural Networks" (2019)

Centenary of Human Population Genetics, Moscow, Russia – Poster: "Ancestry-specific analyses reveal differential demographic histories and opposite selective pressures in modern South Asian populations" (2019)

EMBO | EMBL Symposium: Reconstructing the Human Past – Using Ancient and Modern Genomics, Heidelberg, Germany – Flash talk/poster: "Ancestry-specific analyses reveal differential demographic histories and opposite selective pressures in modern South Asian populations" (2019)

Publications

Yelmen, B., Marnetto, D., Molinaro, L., Flores, R., Mondal, M. and Pagani, L., 2021. Improving selection detection with population branch statistic on admixed populations. *Genome Biology and Evolution*.

Molinaro, L., Marnetto, D., Mondal, M., Ongaro, L., Yelmen, B., Lawson, D.J., Montinaro, F. and Pagani, L., 2021. A Chromosome-Painting Based Pipeline to Infer Local Ancestry under Limited Source Availability. *Genome Biology and Evolution*.

Yelmen, B., Decelle, A., Ongaro, L., Marnetto, D., Tallec, C., Montinaro, F., Furtlehner, C., Pagani, L. and Jay, F., 2021. Creating artificial human genomes using generative neural networks. *PLoS genetics*.

Montinaro, F., Pankratov, V., Yelmen, B., Pagani, L. and Mondal, M., 2020. Revisiting the Out of Africa event with a novel Deep Learning approach. *bioRxiv*. (submitted preprint)

Çiftçi, A., Yelmen, B., Değirmenci, F.Ö. and Kaya, Z., 2020. Impact of biased sex ratio on the genetic diversity, structure, and differentiation of Populus nigra (European black poplar). *Botany*.

Molinaro, L., Montinaro, F., Yelmen, B., Marnetto, D., Behar, D.M., Kivisild, T. and Pagani, L., 2019. West Asian sources of the Eurasian component in Ethiopians: a reassessment. *Scientific reports*.

Yelmen, B., Mondal, M., Marnetto, D., Pathak, A.K., Montinaro, F., Gallego Romero, I., Kivisild, T., Metspalu, M. and Pagani, L., 2019. Ancestry-specific analyses reveal differential demographic histories and opposite selective pressures in modern South Asian populations. *Molecular biology and evolution*.

Extracurricular Activities

Photography and videography

Music and sound production

ELULOOKIRJELDUS

Sünnikuupäev: 18. märts 1990
Sünnikoht: İzmir, Türgi
Kodakondsus: Türgi

Hariduskäik
Tartu Ülikool, molekulaar- ja rakubioloogia instituut, geenitehnoloogia, **doktoriõpe** (2017–…)
Lähis-Ida Tehnikaülikool, eluteaduste õppetool, **magistriõpe** (2014–2017)
Lähis-Ida Tehnikaülikool, molekulaarbioloogia ja geneetika õppetool, **bakalaureuseõpe** (2008–2014)
Ankara Atatürk Anadolu Keskkool (2004–2008)

Töökogemus
Tartu Ülikool, genoomika instituut, **moodsa populatsioonigeneetika nooremteadur** (2017–…)
IQ Smart Software Solutions, **programmeerija** (2016)
TUBITAK (Türgi Teadusliku ja Tehnoloogilise Uurimise Nõukogu) uurimisprojektid, Paju liikide molekulaarne fülogeneetika ja majanduslikult oluliste liikide (*Salix alba* ja *Salix excelsa*) geeniressursside iseloomustamine parendamise eesmärgil, **stipendiaat** (2014–2015)
INTERGEN Geneetika ja Harvikhaiguste Diagnostika, teadus- ja arenduskeskus – **praktikant** (2013)

Õpetamiskogemus
Tartu Ülikool, evolutsioonibioloogia õppetool, Elusloodus ja evolutsioon – (2020/2021 sügissemester)
Tartu Ülikool, evolutsioonibioloogia õppetool, Elusloodus ja evolutsioon – (2019/2020 sügissemester)
Tartu Ülikool, evolutsioonibioloogia õppetool, Evolutsiooniprotsessid – (2018/2019 kevadsemester)
Tartu Ülikool, evolutsioonibioloogia õppetool, Elusloodus ja evolutsioon – (2018/2019 sügissemester)
Tartu Ülikool, evolutsioonibioloogia õppetool, Evolutsiooniprotsessid – (2017/2018 kevadsemester)

Juhendamine
Gleb Kovalev (B.Sc.) – **lõputöö juhendaja** (2020–2021)
Annabella Kadavanich (B.Sc.) – **praktikandi juhendaja** (2020)

Keeleoskus
türgi (emakeel)
inglise (vabalt valdamine)

Konverentsiettekanded ja postrid

ESHG Euroopa inimgeneetika virtuaalne konverents – interaktiivne e-poster: "Kunstlike inimgenoomide loomine generatiivsete närvivõrkude abil" (2020)

MRI ja GI aastakonverents, Tartu, Eesti – ettekanne: "Kunstlike inimgenoomide loomine generatiivsete närvivõrkude abil" (2019)

100 aastat inimese populatsioonigeneetikat, Moskva, Venemaa – poster: "Põlvnemisspetsiifilistest analüüsidest selguvad tänapäeva Lõuna-Aasia populatsioonide erinevad demograafilised ajalood ja vastandlikud valikusurved" (2019)

EMBO | EMBL Sümpoosium: Inimese mineviku taastamine – kasutades iidset ja moodsat genoomikat, Heidelberg, Saksamaa – välkettekanne/ poster: "Põlvnemisspetsiifilistest analüüsidest selguvad tänapäeva Lõuna-Aasia populatsioonide erinevad demograafilised ajalood ja vastandlikud valikusurved" (2019)

Publikatsioonid

Yelmen, B., Marnetto, D., Molinaro, L., Flores, R., Mondal, M. ja Pagani, L., 2021. Valiku tuvastamise parandamine segunenud populatsioonides populatsiooni haru statistiku abil. *Genome Biology and Evolution*.

Molinaro, L., Marnetto, D., Mondal, M., Ongaro, L., Yelmen, B., Lawson, D.J., Montinaro, F. ja Pagani, L., 2021. Kromosoomide "värvimisel" põhinev toru lokaalse põlvnemise tuletamiseks allikate piiratud kättesaadavuse tingimustes. *Genome Biology and Evolution*.

Yelmen, B., Decelle, A., Ongaro, L., Marnetto, D., Tallec, C., Montinaro, F., Furtlehner, C., Pagani, L. ja Jay, F., 2021. Kunstlike inimgenoomide loomine generatiivsete närvivõrkude abil. *PLoS genetics*.

Montinaro, F., Pankratov, V., Yelmen, B., Pagani, L. ja Mondal, M., 2020. Aafrikast välja rändamise sündmuse uus analüüs uudse süvaõppimise lähenemisega. *bioRxiv*. (submitted preprint)

Çiftçi, A., Yelmen, B., Değirmenci, F.Ö. ja Kaya, Z., 2020. Kallutatud sugude suhte mõju *Populus nigra* (Euroopa must pappel) geneetilisele varieeruvusele, struktuurile ja diferentseerumisele. *Botany*.

Molinaro, L., Montinaro, F., Yelmen, B., Marnetto, D., Behar, D.M., Kivisild, T. ja Pagani, L., 2019. Etiooplaste Euraasia komponendi Lääne-Aasia allikad: uus hinnang. *Scientific reports*.

Yelmen, B., Mondal, M., Marnetto, D., Pathak, A.K., Montinaro, F., Gallego Romero, I., Kivisild, T., Metspalu, M. ja Pagani, L., 2019. Põlvnemisspetsiifilistest analüüsidest selguvad tänapäeva Lõuna-Aasia populatsioonide erinevad demograafilised ajalood ja vastandlikud valikusurved. *Molecular biology and evolution*.

Muud tegevused

Fotograafia ja videograafia
Muusika ja heliproduktsioon

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

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

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

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

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

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

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

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

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

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

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

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

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

246. **Mario Plaas**. Animal model of Wolfram Syndrome in mice: behavioural, biochemical and psychopharmacological characterization. Tartu, 2013, 144 p.
247. **Georgi Hudjašov**. Maps of mitochondrial DNA, Y-chromosome and tyrosinase variation in Eurasian and Oceanian populations. Tartu, 2013, 115 p.
248. **Mari Lepik**. Plasticity to light in herbaceous plants and its importance for community structure and diversity. Tartu, 2013, 102 p.
249. **Ede Leppik**. Diversity of lichens in semi-natural habitats of Estonia. Tartu, 2013, 151 p.
250. **Ülle Saks**. Arbuscular mycorrhizal fungal diversity patterns in 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.
361. **Erwan Pennarun**. Meandering along the mtDNA phylogeny; causerie and digression about what it can tell us about human migrations. Tartu, 2019, 162 p.
362. **Karin Ernits**. Levansucrase Lsc3 and endo-levanase BT1760: characterization and application for the synthesis of novel prebiotics. Tartu, 2019, 217 p.
363. **Sille Holm**. Comparative ecology of geometrid moths: in search of contrasts between a temperate and a tropical forest. Tartu, 2019, 135 p.

364. **Anne-Mai Ilumäe**. Genetic history of the Uralic-speaking peoples as seen through the paternal haplogroup N and autosomal variation of northern Eurasians. Tartu, 2019, 172 p.
365. **Anu Lepik**. Plant competitive behaviour: relationships with functional traits and soil processes. Tartu, 2019, 152 p.
366. **Kunter Tätte**. Towards an integrated view of escape decisions in birds under variable levels of predation risk. Tartu, 2020, 172 p.
367. **Kaarin Parts**. The impact of climate change on fine roots and root-associated microbial communities in birch and spruce forests. Tartu, 2020, 143 p.
368. **Viktorija Kukuškina**. Understanding the mechanisms of endometrial receptivity through integration of 'omics' data layers. Tartu, 2020, 169 p.
369. **Martti Vasar**. Developing a bioinformatics pipeline gDAT to analyse arbuscular mycorrhizal fungal communities using sequence data from different marker regions. Tartu, 2020, 193 p.
370. **Ott Kangur**. Nocturnal water relations and predawn water potential disequilibrium in temperate deciduous tree species. Tartu, 2020, 126 p.
371. **Helen Post**. Overview of the phylogeny and phylogeography of the Y-chromosomal haplogroup N in northern Eurasia and case studies of two linguistically exceptional populations of Europe – Hungarians and Kalmyks. Tartu, 2020, 143 p.
372. **Kristi Krebs**. Exploring the genetics of adverse events in pharmacotherapy using Biobanks and Electronic Health Records. Tartu, 2020, 151 p.
373. **Kärt Ukkivi**. Mutagenic effect of transcription and transcription-coupled repair factors in *Pseudomonas putida.* Tartu, 2020, 154 p.
374. **Elin Soomets**. Focal species in wetland restoration. Tartu, 2020, 137 p.
375. **Kadi Tilk**. Signals and responses of ColRS two-component system in *Pseudomonas putida.* Tartu, 2020, 133 p.
376. **Indrek Teino**. Studies on aryl hydrocarbon receptor in the mouse granulosa cell model. Tartu, 2020, 139 p.
377. **Maarja Vaikre**. The impact of forest drainage on macroinvertebrates and amphibians in small waterbodies and opportunities for cost-effective mitigation. Tartu, 2020, 132 p.
378. **Siim-Kaarel Sepp**. Soil eukaryotic community responses to land use and host identity. Tartu, 2020, 222 p.
379. **Eveli Otsing**. Tree species effects on fungal richness and community structure. Tartu, 2020, 152 p.
380. **Mari Pent**. Bacterial communities associated with fungal fruitbodies. Tartu, 2020, 144 p.
381. **Einar Kärgenberg**. Movement patterns of lithophilous migratory fish in free-flowing and fragmented rivers. Tartu, 2020, 167 p.
382. **Antti Matvere**. The studies on aryl hydrocarbon receptor in murine granulosa cells and human embryonic stem cells. Tartu, 2021, 163 p.
383. **Jhonny Capichoni Massante**. Phylogenetic structure of plant communities along environmental gradients: a macroecological and evolutionary approach. Tartu, 2021, 144 p.

384. **Ajai Kumar Pathak.** Delineating genetic ancestries of people of the Indus Valley, Parsis, Indian Jews and Tharu tribe. Tartu, 2021, 197 p.
385. **Tanel Vahter.** Arbuscular mycorrhizal fungal biodiversity for sustainable agroecosystems. Tartu, 2021, 191 p.