DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
384

AJAI KUMAR PATHAK

Delineating genetic ancestries of people of the Indus Valley, Parsis, Indian Jews and Tharu tribe

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
384

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
384

AJAI KUMAR PATHAK

Delineating genetic ancestries of people of the Indus Valley, Parsis, Indian Jews and Tharu tribe

UNIVERSITY OF TARTU
Press

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

Dissertation was accepted for the commencement of the degree of Doctor of Philosophy in Molecular and Cell Biology on 3rd of May 2021 by the Council of the Institute of Molecular and Cell Biology, Faculty of Science and Technology, University of Tartu.

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

Professor Gyaneshwer Chaubey
Cytogenetics laboratory,
Department of Zoology
Banaras Hindu University

Opponent: Professor Francesc Calafell
Departament de Ciències Experimentals i de la Salut
Universitat Pompeu Fabra
Institut de Biologia Evolutiva (CSIC-UPF)

Commencement: Room No 105, 23B Riia St, Tartu, on 15th of June 2021, at 14:15.

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

ISSN 1024-6479
ISBN 978-9949-03-608-0 (print)
ISBN 978-9949-03-609-7 (pdf)

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF FIGURES 8

LIST OF ORIGINAL PUBLICATIONS 9

ABBREVIATIONS 11

1. INTRODUCTION 12

2. LITERATURE OVERVIEW 15

2.1. The perspective of genetic variation in human evolutionary history 15
2.1.1. Classical protein markers 15
2.1.2. Haploid parts of genome 16
2.1.2.1. Mitochondrial DNA 16
2.1.2.2. Y chromosome 18
2.1.3. Whole genomes 20

2.2. The power of aDNA to decipher the human demographic history 22
2.2.1. aDNA revealing archaic admixture in modern humans 24

2.3. Methods to detect genome-wide patterns of population structure and demographic history 26
2.3.1. Mean pairwise genetic distance (F_{ST}) 26
2.3.2. Spatial techniques to infer ancestry at a global level 26
2.3.3. Clustering algorithms-based methods 27
2.3.4. Inferring a fine-grained picture of admixture history from haplotypes 29
2.3.4.1. Approaches of local ancestry inference 29
2.3.4.2. Approaches that side-step local ancestry inference 30
2.3.4.3. Identical-by-descent (IBD) approach 32
2.3.5. Dating admixture using pattern of LD decay 33
2.3.6. Identifying admixture using drift parameters 35
2.3.7. Tree-based tools of population splits and admixture 38

2.4. South Asia in general 40
2.4.1. Salient features of South Asia 42
2.4.1.1. Climate and Landscapes 43
2.4.1.2. Languages 43
2.4.1.3. The Social and Caste system 44
2.4.2. Northwest India: the root of ancient Harappan Culture and Vedic Kurukshetra 46
2.4.2.1. Ancient Harappa 46
2.4.2.2. Vedic India and ancient Kurukshetra 50

2.5. Archaeological attributes of South Asia 51
2.5.1. Evidence from fossil records 52
2.5.2. Evidence from Paleolithic and Microlithic tools 53

2.6. Genetic overview of South Asia 55
2.6.1. Demographic inferences from haploid data 55
2.6.2. Demographic inferences of South Asia from genome-wide data 56
2.6.3. Demographic inferences of South Asia from aDNA 57
2.6.3.1. Pre- and early historical demography of South Asia . 57
2.6.3.2. Archaic admixture pattern of South Asia 60
2.6.4. Languages and genetic correlation in South Asia 61
2.6.5. Sex-specific ancestry patterns 62
2.6.6. Historical immigrants into South Asia and their genetic antiquity 63
2.7. Parsis and Jews: the case of two recent immigrants from the Near and Middle East 64
2.7.1. Parsis 64
2.7.2. Jews 65
2.8. Particular case of linguistically diverse Tharu tribe from Tarai region 66
3. OBJECTIVES OF THE STUDY 68
4. MATERIALS AND METHODS 69
5. RESULTS AND DISCUSSION 70
5.1. Genetic variability among modern Indus Valley populations suggest a population-specific demographic history and ancient links (Ref I) 70
5.1.1. Genetic ancestry composition of populations from east and west Indus basin 70
5.1.2. Modern Indus Valley populations in terms of ancient West Eurasian ancestries 72
5.1.3. Unique genetic structure of Ror population 73
5.1.3.1. Alternate ANI representative 74
5.1.4. Recent updates from the first aDNA of Indus Valley *compared to* our study 74
5.2. Genetic outline of historical migrants to South Asia: The case of Parsi and Jewish populations (Ref II/ Ref III) 75
5.2.1. Population structure and demographic history of Parsis 75
5.2.1.1. Genome-wide ancestry of Indian and Pakistani Parsis 76
In the context of modern populations 76
In the context of ancient individuals 77
5.2.1.2. Maternal and paternal lineage of Parsis 78
5.2.2. Jews in the context of ancient West Eurasian individuals 79

5.3. Reconstructing the genetic history of Tharu people:
a linguistically diverse tribal group of Tarai region (Ref IV) 80
5.3.1. Genome-wide variation of the Tharu groups 80
5.3.2. Maternal and paternal lineage diversity in the Tharu 81

6. CONCLUSIONS 84

SUMMARY IN ESTONIAN 85

REFERENCES 89

ACKNOWLEDGEMENTS 120

PUBLICATIONS 123

CURRICULUM VITAE 172

ELULOOKIRJELDUS 176

LIST OF FIGURES

Figure 1. PCA and STRUCTURE like analyses in simulated data. 28
Figure 2. Inferring fine-scale ancestry from simulated data..... 31
Figure 3. Dating admixture applying ROLLOFF and GLOBETROTTER to simulated data. 34
Figure 4. Representation of f- and D-statistics. 36
Figure 5. Inferred human phylogenetic tree fitting admixture events. 39
Figure 6. Early and later modern human dispersal routes during Late Pleistocene. 42
Figure 7. Language map of India. 44
Figure 8. Map of ancient Indus Valley settlements. 48
Figure 9. A sketch of the peopling history of South Asia. 59
Figure 10. IBD Sharing and Ancestry Profile of modern Indus Valley Populations. 71
Figure 11. Population structure and ancient ancestry profile of Parsi and Jews. 77
Figure 12. Haplotype sharing and haploid genetic diversity among Tharu and neighboring groups. 82

LIST OF ORIGINAL PUBLICATIONS

Ref I

Pathak, Ajai K; Kadian, Anurag; Kushniarevich, Alena; Montinaro, Francesco; Mondal, Mayukh; Ongaro, Linda; Singh, Manvendra; Kumar, Pramod; Rai, Niraj; Parik, Jüri; Metspalu, Ene; Rootsi, Siiri; Pagani, Luca; Kivisild, Toomas; Metspalu, Mait; Chaubey, Gyaneshwer; Villems, Richard (2018).
The Genetic Ancestry of Modern Indus Valley Populations from Northwest India. American Journal of Human Genetics, 103 (6), 918–929.
https://doi.org/10.1016/j.ajhg.2018.10.022.

Ref II

Chaubey, Gyaneshwer; Ayub, Qasim; Rai, Niraj; Prakash, Satya; Mushrif-Tripathy, Veena; Mezzavilla, Massimo; **Pathak, Ajai K**; Tamang, Rakesh; Firasat, Sadaf; Reidla, Maere; Karmin, Monika; Rani, Deepa Selvi; Reddy, Alla G; Parik, Jüri; Metspalu, Ene; Rootsi, Siiri; Dalal, Kurush; Khaliq, Shagufta; Mehdi, Syed Qasim; Singh, Lalji; Metspalu, Mait; Kivisild, Toomas; Tyler-Smith, Chris; Villems, Richard; Thangaraj, Kumarasamy (2017).
"Like sugar in milk": reconstructing the genetic history of the Parsi population. Genome Biology, 18 (1), 110–110.
https://doi.org/10.1186/s13059-017-1244-9.

Ref III

Pathak, Ajai K; Srivastava, Anshika; Singh, Prajjval Pratap; Das, Debashruti; Bandopadhyay, Audditiya; Singh, Pooja; Chaubey, Gyaneshwer (2019).
Historic migration to South Asia in the last two millennia: A case of Jewish and Parsi populations. Journal of Biosciences, 44 (3), UNSP 72.
https://doi.org/10.1007/s12038-019-9873-7.

Ref IV

Chaubey, Gyaneshwer; Singh, Manvendra.; Crivellaro, Federica; Tamang, Rakesh; Nandan, Amrita.; Singh, Kamayani; Sharma, Varun K; **Pathak, Ajai K**; Shah, Anish M; Sharma, Vishwas; Singh, Vipin K; Selvi Rani, D; Rai, Niraj; Kushniarevich, Alena; Ilumäe, Anne-Mai; Karmin, Monika; Phillip, Anand; Verma, Abhilasha; Prank, Erik; Singh, Vijay K; Li, Blaise; Giovindaraj, Periyasami; Chaubey, Akhilesh K; Dubey, Pavan K; Reddy, Alla G; Premkumar, Kumpati; Vishnupriya, Satti; Pande, Veena; Parik, Jüri; Rootsi, Siiri; Endicott, Phillip; Metspalu, Mait; Lahr, Marta M; Van Driem, G; Villems, Richard; Kivisild, Toomas; Singh, Lalji; Thangaraj, K. (2014).
Unravelling the distinct strains of Tharu ancestry. European Journal of Human Genetics, 22 (12), 1404–1412. https://doi.org/10.1038/ejhg.2014.36.

*For complete list of research papers, co-authored by **Ajai K. Pathak**., see Curriculum Vitae (English).*

Author's contribution to the listed publications are mentioned as bellow:

Ref I – had a role in the study design, performed lab experiments and statistical analysis of the data, interpreted results, and wrote the paper.

Ref II – performed the haploid DNA genotyping and analyses thereof, and assisted in manuscript preparation.

Ref III – performed the statistical analyses, interpreted results, and wrote the paper with the help of co-authors.

Ref IV – participated in data analysis, interpretation of results, and the preparation of the article.

ABBREVIATIONS

AASI	Ancient Ancestral South Indians
aDNA	ancient DNA
AMH	anatomically modern humans
Anatolia_N	Anatolian Neolithic
ANI	Ancestral North Indian
ASI	Ancestral South Indian
Austroasiatic	AA
B/CE	before common era
BA	Bronze Age
CHG	Caucasus hunter-gatherers
chrY	Y chromosome
EDAR	Ectodysplasin A receptor
EHG	Eastern hunter-gatherers
EMBA	Early Middle Bronze Age
F_{ST}	fixation index, mean pairwise genetic distance
H. sapiens	*Homo sapiens*
Hg	haplogroup
HMM	Hidden Markov model
IBD	identical by descent
IN_IE	Indian Indo-European
IP	Indus Periphery
Iran_N	Iranian Neolithic
IVC	Indus Valley Civilization
KYA	thousand years ago
LD	Linkage disequilibrium
MAF	minor allele frequency
MCMC	Markov chain Monte Carlo
MLBA	Middle to late Bronze Age
MSY/NRY	male-specific regions/non-recombining part of chrY
mtDNA	mitochondrial DNA
Ne	effective population size
NGS	next-generation sequencing
NNLS	non-negative least square
NWI	Northwest India
OoA	Out of Africa
PAR	pseudo autosomal region of chrY
PCA	principal component analysis
SE Asia	Southeast Asia
SNP	single nucleotide polymorphism
SRY	sex-determining region of chrY
STR	short tandem repeats
TB	Tibeto-Burmans
WS_HG	Western Siberian Hunter-Gatherers

1. INTRODUCTION

Though curiosity is an inherent feature of most animal species, it is often considered a characteristic feature of the great apes and humans. This thirst for knowledge has been the driving force in our search for when, where, and how the human species originated and evolved. Since the split from other great apes, humans have gone through several evolutionary changes in their phenotypes and genomes. Experts from different research fields have been studying the phenotypic, cultural, and genetic variations to get an insightful inference of the human population history, sparking the general public's interest. The studies that attempt to trace the demographic history of a population use several kinds of genetic data, extending from the classical (i.e., polymorphic blood proteins) and haploid markers – mitochondrial DNA (mtDNA) and Y chromosome (chrY) – to whole genomes. The study of uniparentally inherited mtDNA and chrY variation offers a limited evolutionary perspective on a population's history, which, however, can reveal sex-biased processes. While most of the nuclear genome is diploid and recombinationally active, its study can provide a broad spectrum view on a population/species' evolutionary history. Thus, applying advanced statistical tools on the genome-wide single nucleotide polymorphism (SNP) genotyping or sequencing data, containing thousands of times more information from multiple loci than the single locus mtDNA and chrY, is a widely preferred approach for population genetic studies. Recent advances in the recovery and analyses of ancient DNA (aDNA) have supplemented human demographic history studies to the extent that seemed impossible only with modern genomes.

Irrespective of the consensus for an African origin of Anatomically Modern Humans (AMH), i.e., *Homo sapiens* (*H. sapiens*), a specific population and/or region of Africa could not be linked to the evolution of AMH. Instead, the growing shreds of evidence suggest AMH evolved within a group of interconnected populations living across Africa (Scerri et al., 2018). Recently found archaeological evidence indicates the origin of AMH in West Africa ~315 thousand years ago (KYA) (Hublin et al., 2017). However, the individual still has a few primitive morphological features. The first established thoroughly modern human fossil was discovered in Ethiopia dated ~195 KYA (McDougall et al., 2005). Some earlier studies suggested the AMH emerged rapidly around 200 KYA among early *H. sapiens* (Stringer, 2002), while others endorse a gradual evolution over the last 400,000 years (Bräuer, 2008). Recent studies analyzing whole human genomes from across the world (Malaspinas et al., 2016; Mallick et al., 2016; Mondal et al., 2016; Pagani et al., 2016) have verified numerous earlier archaeological and genetic studies about the departure of modern humans from Africa around 70–52 KYA. Most of the contemporary peoples outside Africa descend from this main Out of Africa (OoA) expansion; however, they conceivably received some (though limited) genetic contribution from earlier Out of Africa (xOoA) expansion(s) (e.g., Pagani et al., 2016). Genetic hints indicate modern human admixture with Neanderthals earlier than the OoA expansion (Kuhlwilm

et al., 2016; Prüfer et al., 2017). Which, together with the fossil and archaeological evidence of *H. sapiens* colonizing Greece and Levant (200–185 KYA), Arabia (~85 KYA), China (~80 KYA), Southeast Asia (73–63 KYA), and Australia (~65 KYA) (Liu et al., 2015; Clarkson et al., 2017; Westaway et al., 2017; Groucutt et al., 2018; Hershkovitz et al., 2018a; Harvati et al., 2019) are in concordance with xOaA. The geographical location and archaeological relics of Middle Paleolithic hominins, along with the robust genetic connection between Australian Aborigines and modern South Asians, point to the likely significance of South Asia in the early modern human dispersals (Rasmussen et al., 2011; Reich et al., 2011; Mondal et al., 2018). The evidence of Archaic introgression from Denisovans to South Asian populations also corroborates the claim mentioned above (Reich et al., 2011; Vernot and Pääbo, 2018; Jacobs et al., 2019). On the other hand, the advent of agriculture ~10 KYA introduced substantial changes in economic and cultural practices – the transition from foraging practices to farming and the spread of languages, likely caused by the movement of peoples related to Neolithic farmers and Bronze Age Steppe herders among others (Haak et al., 2015; Lazaridis et al., 2016; Narasimhan et al., 2019). Such movements also contributed significantly to the ancestry of contemporary South Asians, probably through the early Indus Valley Civilization (IVC) peoples in the northwest region of the Indian subcontinent (Majumder and Basu, 2015; Narasimhan et al., 2019). Serving as a gallery for human movements during pre- and early historical times, South Asia warrants a greater insight into the genetic architecture of inadequately represented regional groups to decipher its demographic history.

This thesis brings up additional knowledge about South Asia's demographic dynamics, analyzing different units of modern genomes, yet focused on the genome-wide SNPs. The dissertation studies the genetic variation of contemporary peoples in the historical Indus valley region of Northwest India (NWI), historical immigrants from the Middle East, and a linguistically heterogeneous tribe of the northern Indian Tarai region. First and foremost, the thesis aims to refine the underrepresented genetic structure of modern Indus Valley people by generating and analyzing new data from the northwest region's contemporary populations (Ref I). The study looks mainly into the genome-wide variation of modern populations in the context of current global people and available ancient individuals, revealing fine-scale population structure and genetic heterogeneity among the modern Indus populations. The study also uncovered a population-specific demographic processes for culturally distinct groups of the region. Secondly, the thesis includes South Asian Parsi and Jewish communities – early historic migrants from the Middle East – aiming to get insight into their genetic structure in the context of modern and ancient individuals and how they assimilated with local peoples (Ref II/III). Ref II analyzes uniparental and autosomal data to pinpoint the genetic homogeneity between different groups of South Asian Parsis, population inbreeding, and arrival time to South Asia. Ref III studies the Parsi and Jewish groups in-depth by reconstructing the broader genetic profile of both the ethnic groups, especially in the context of the available ancient genomes.

Finally, the dissertation focuses on the Tharu community, a linguistically diverse tribal group from the Tarai region of northern India, to look into the origin and intra-/inter-population genetic variation relative to their neighbors (Ref IV). The genome-wide data discloses a genetic heterogeneity between Tharu groups followed by a substantial variation in maternal and paternal lineages, suggesting differences in the founding lineages and successive assimilation of Tharu groups. Thus, the study supports a model of cultural rather than demic unity of Tharu groups, and their likely origin in the northern region of the subcontinent.

2. LITERATURE OVERVIEW

The literature overview firstly recapitulates the essential features of different types of genomic data and the significance of ancient DNA (aDNA) in decoding the population's genetic history. Later chapters summarize the key methods employed as tools in population genetic studies before outlining South Asia's social, linguistic, archaeological and genetic synopsis pertinent to the specific questions addressed in this dissertation.

2.1. The perspective of genetic variation in human evolutionary history

Populations are generally assumed to be the pool of homogeneous genes where individuals mate randomly with each other. Nevertheless, natural populations are spread over larger areas, separated by different geographical barriers and environmental conditions, and have experienced several evolutionary changes through time, resulting in variation between populations. Indeed, several demographic factors, such as scarcity of natural resources due to the increasing population size, instigate populations to migrate to new geographical regions for survival, forming new subpopulations. Individuals of subpopulations may interact and mate with those living in the same local habitat more often; such random fluctuation in mating leads to changes in frequencies of different alleles (genetic drift). Genetic drift makes individuals of a subpopulation genetically similar while causing the difference among subpopulations.

Additionally, natural selection – the adaptation of an organism to the environment by inheriting the variation that increases the survival and reproduction of its individuals- may also contribute to the genetic variation of populations. Despite the differentiation of people caused by whatever demographic processes, they carry a "universal" genomic feature that can be traced back to the common ancestor. Overall, in the absence of consistent admixture, the genetic distance between any two populations positively correlates with their split time from each other.

Thus, the study of the accumulation of genetic differences among populations can reconstruct population histories. Genetic studies that attempt to trace the population history use several kinds of data, ranging from classical protein markers and uniparentally inherited parts of the genome to the sequences of whole genomes. Here, I briefly summarize their use in population genetic studies.

2.1.1. Classical protein markers

The first set of the classical markers used to infer the human population histories were ABO blood groups and different classes of proteins (Cavalli-Sforza et al., 1994). The ABO blood group system was the first defined genetic variation

(polymorphism) of humans (Landsteiner, 1900). The initial population genetic studies leveraged the frequency difference of ABO blood groups in different populations. The other blood groups (e.g., MN and Rh) independent of the ABO system were defined and used for the same purposes later (Levine and Stetson, 1939; Landsteiner and Wiener, 1940). The introduction of protein separation methods like electrophoresis enabled more classical markers – variants of immunoglobulins and highly polymorphic leucocyte antigens (HLA), to study the genetic diversity of human populations (Cavalli-Sforza et al., 1994). Some highlights of that period include the observed pattern of reduced genetic diversity in non-Africans and Near East to Northwest European genetic cline, correlating well with the hypothesis of out-of-Africa migration and Neolithic dispersal of agriculturists, respectively. In the case of South Asia too, the classical protein markers have remarkably been used in the early and later population genetic studies (Chahal and Papiha, 1981; Papiha et al., 1982; Mastana and Papiha, 1992; Mohyuddin and Mehdi, 2005; Raza et al., 2013; Mohsenpour and Chandrasekar, 2019).

2.1.2. Haploid parts of genome

Haploid parts of the genome, often used in population genetic studies since the early 1990-ies, are mitochondrial (mt) and Y chromosomal (chrY) DNA, also known as uniparental loci. Unlike the rest of the genome, these markers are unique in their mode of inheritance, which is beneficial in tracking the ancestral lineages in a sex-specific manner; chrY for paternal lineages and mtDNA for maternal lineages. A relatively high mutation rate is a force behind the rapid changes in mtDNA and chrY sequences. These mutations characterize the differences between individual lines, and the number of accumulated nucleotide changes is suggestive of the time in the past when two lineages had a common ancestor.

2.1.2.1. Mitochondrial DNA

The human genome has two sovereign genetic components, the complex nuclear genome encompassing most (~99.999%) of the genetic information on an individual. In contrast, the "secondary" genome or the cytoplasmic genome of mitochondria accounts for an individual's minor but vital genetic information. The circular mtDNA is a double-stranded molecule of ca 16,5 kbp (Andrews et al., 1999). The presence of mitochondria – the powerhouse of our cells that produce energy – is a unique feature of eukaryotes and has been phenomenal in the evolution of eukaryotic cells (Kurland et al., 2006; Kivisild, 2015). mtDNA has several features, such as the lack of recombination (Merriwether et al., 1991; Hagström et al., 2014), maternal inheritance (Hutchison et al., 1974), high copy number per cell (Michaels et al., 1982), and fast mutation rate (Brown et al.,

1979). Altogether, these features had attracted researchers to use it as a genetic tool to study human evolutionary history, especially before human genome sequencing was achieved. A few studies claimed a likely transmission of paternal mtDNA to the next generation (Schwartz and Vissing, 2002; Luo et al., 2018). Nonetheless, further investigations raised doubt and could not find adequate evidence for the paternal inheritance of mtDNA (Filosto et al., 2003; Taylor et al., 2003; Schwartz and Vissing, 2004; Luo et al., 2019; Lutz-Bonengel and Parson, 2019). Additional studies have refuted the occurrence of recombination between maternal and paternal mtDNA because even if it is leaked, the paternal mtDNA found in the zygote is eliminated by a mechanism functional at a molecular level mostly (Carelli, 2015; Pyle et al., 2015).

The non-recombing mtDNA contains 37 genes, out of which 13 encode proteins, while the rest 24 genes that comprise two ribosomal RNAs (rRNAs) and 22 transfer RNAs (tRNAs) assist in the translation of those 13 polypeptides (Anderson et al., 1981). The primary function of most of these genes is to produce energy and store it in the form of ATP, the cell's principal source of energy. mtDNA lacks the proofreading mechanism though they have a few prominent DNA repair mechanisms (Pinz and Bogenhagen, 1998; Mason et al., 2003; Szczesny et al., 2008; Rossi et al., 2009), these are inadequate to correct the over-burden mtDNA damages caused by the oxidative stress (Tuppen et al., 2010). Oxidative stress results from the surge in the production of reactive oxygen species referred to as ROS (Mikhed et al., 2015), one of the leading sources for mtDNA mutations that adds up with elderly and in disease conditions. ROS are formed by the electron transport system (ETS) of mitochondria (Shokolenko et al., 2009). mtDNA evolves faster than the nuclear DNA, as mutations are passed on to the next generation in the absence of repair and recombination. The data from coding and non-coding parts of mtDNA can be pooled to make a phylogenetic tree. The expanding branches of the phylogenetic tree might be characterized by using a set of restriction fragment length polymorphisms (RFLPs) specific to a particular branch, resulting in the mtDNA haplogroups, i.e., most common branches assigned by alphabetical labels (Torroni et al., 1993). On average, the mtDNA mutation rate in animals is an order of magnitude higher than that of nuclear DNA (Brown et al., 1979). It is even higher in the non-coding control region than the coding region. The non-coding region of mtDNA consists of three hypervariable segments, HVSI, HVSII, and HVSIII. The higher nucleotide polymorphisms in the three HVS regions, especially HVSI and HVSII, accumulate mutations almost ten folds faster than other mtDNA regions (Soares et al., 2009). Despite the low resolution, the hypervariable control region was extensively used in the early human population genetic studies (Torroni et al., 1993, 1998; Richards et al., 1996). Later studies started using complete mtDNA sequences to reconstruct the human mtDNA phylogeny (Torroni et al., 2006; Behar et al., 2008b; Van Oven and Kayser, 2009). The whole-genome sequence of mtDNA can be applied as a high molecular resolution tool to identify distinct genetic patterns accumulated in the human population over thousands of years (Kivisild, 2015). In one such attempt, Behar et al. (Behar et al., 2008b) applied whole mtDNA

sequences reporting the early human settlements in Africa displaying structured mtDNA ancestries. The mtDNA lineage of South African KhoiSan people diverged early (90–150 KYA) from the rest of the human mtDNA. A similar pattern was replicated by a recent mitogenomes study (Chan et al., 2019).

However, the classification of human mtDNA phylogeny remained clumsy as mutations are identified against the rCRS reference, the first wholly sequenced mtDNA of a present-day European (Anderson et al., 1981; Andrews et al., 1999). Nevertheless, a landmark study (Behar et al., 2012) presented the Reconstructed Sapiens Reference Sequence (RSRS) as a reference with most parsimonious updates to the previously proposed human mtDNA root by incorporating available mitogenomes of *Homo neanderthalensis*.

In brief, though the low effective population size (one quarter to that of any autosome) makes mtDNA predisposed to genetic drift than autosomes, its attributes in lacking the recombination and inheritance on maternal line render mtDNA a fantastic tool to study the female aspect of population history.

2.1.2.2. Y chromosome

The nuclear genetic material of a human somatic cell consists of 46 separate chromosomes, divided into 23 pairs; one of each pair is passed from both parents to offspring. Twenty-two pairs (1–22) are identical in males and females and are referred to as autosomes. At the same time, the remaining two chromosomes differ between females (two copies of the X chromosome (chrX) and males (having one chrY in addition to one chrX), therefore, termed as the sex chromosomes. ChrY chromosome is male-specific, haploid, and inherited only paternally from father to son; it is the smallest one among all chromosomes that also carry the smallest number of genes of a chromosome (Jobling and Tyler-Smith, 1995).

Most of ~60 Mb of the chrY does not recombine and is referred to as the non-recombining portion (NRY) of the chrY, sometimes also called the male-specific region (MSY) (Skaletsky et al., 2003; Jobling and Tyler-Smith, 2017). Nevertheless, the short-specialized portions of chrY do recombine with the chrX during male meiosis due to their well-preserved identical sequences; such regions are called the pseudo-autosomal regions (PAR) (Skaletsky et al., 2003). There are two short pseudo-autosomal regions, PAR1 located on the tip of the short arm (Yp), and PAR2 is located on the tip of the long arm (Yq). PAR1 and PAR2 are euchromatic (genetically active), while between the PAR1 and PAR2 lies the MSY region – comprising both heterochromatic (genetically inactive) and euchromatic region (Jobling and Tyler-Smith, 2003; Skaletsky et al., 2003). The euchromatic part of MSY has three main classes of sequence (Skaletsky et al., 2003; Jobling and Tyler-Smith, 2017): (1) ~8.6 Mb long X-degenerate (XDG) region, (2) X-transposed region (XTR) of ~3.4 Mb size, and (3) 10.2 Mb size ampliconic regions. XDG region has diverged from the ancestral proto-X chromosome; XTR is transferred from X-chromosome when human lineage diverges from human and chimpanzee's common ancestor (Page et al., 1984), while ampliconic regions

are the duplicated sequences of a chromosome that are highly identical. The high inter-chromosomal similarity of the XTR and intra-chromosomal similarity of ampliconic regions pose challenges to interpreting the resequencing data. Therefore, only a prospective, less complex region with explicitly callable variants that extend around 10 Mb on average is used (Poznik et al., 2013; Karmin et al., 2015). However, some scholars suggest that a significantly longer region of ~21 Mb can be used for resequencing (Helgason et al., 2015).

Furthermore, several studies claimed that an alternative recombination process known as the gene conversion happens on chrY (Rozen et al., 2003; Skaletsky et al., 2003), yet, after considering all this, the fact remains that we can treat NRY or MSY practically as a single locus. Male specificity of chrY is helpful to extract the male side of diversity in populations. Besides, chrY plays a role in sex determination via the phenomenal sex-determining region of Y (SRY) that acts during the early development of mammals and triggers a gonad precursor to turn out to be a testis rather than an ovary (Sekido and Lovell-Badge, 2008; Kashimada and Koopman, 2010). The somatic loss of chrY in older adults has also been linked to the higher risk of Alzheimer's (Dumanski et al., 2016) and cancer mortality (Forsberg et al., 2014).

Population-specific diversity of chrY reveals the social structure characterized by the dominance of men or women in marriage practices that affects the movement of males and females between different groups of society (Heyer et al., 2012). A patrilocal society is dominated by males, where married couples reside with the husband's family; thus, women move to husband's residence; this is reflected by genetically less diverse chrY than mtDNA. On the contrary, in a matrilocal society, males move to live with their wife's family; therefore, population-specific diversity of chrY would be higher than that of mtDNA. Male-specific heritage of chrY has been widely used for forensic DNA analysis and genetic genealogy (Jobling et al., 1997; Calafell and Larmuseau, 2017).

The mutation rate of chrY, though slower than that of mtDNA, is, in general, higher than that of autosomal chromosomes because transmission of chrY occurs through male germline, and sperms undergo many-many more cell cycles than oocytes (Jobling and Tyler-Smith, 2003). Different genetic markers of chrY evolve at different rates; single nucleotide polymorphisms (SNPs) have a mutation rate of 10^{-8} compared to the 10^{-3} of short tandem repeats (STRs), and by combining both, it is possible to scan several time points in past (de Knijff, 2000). SNPs define haplogroups which can be applied to construct a phylogenetic tree (Y Chromosome Consortium, 2002; Jobling and Tyler-Smith, 2017), while STRs, whose number varies in different populations, can be applied to extract the diversity within these haplogroups (de Knijff, 2000).

Though the smaller effective population sizes of both the mtDNA and chrY, i.e., one-fourth, relative to autosomes, make them more prone to genetic drift, both can identify the social structure and likely difference in behaviors of men and women (Jobling and Tyler-Smith, 2017). For example, Bayesian Skyline Plots (BSPs), an approach to get the overview of past population sizes, can be applied to the genome sequences of mtDNA and chrY to compare how the female and

male effective population sizes changed through time. The same approach was applied to global genome data (Karmin et al., 2015), which provided a different pattern for the effective population sizes (*Ne*) of males and females. The study documents a noticeable drop in the *Ne* for chrY just around mid-Holocene (8–4 KYA) followed by a subsequent expansion in *Ne* of males. The females displayed a consistently high effective population size than males, whereas the sharp increase of effective population size in women was well pronounced earlier, between 20 and 10 KYA. The study also proposed that the inferred bottleneck of chrY in the last 10 KYA is primarily the consequence of cultural changes that affected male reproductive success.

2.1.3. Whole genomes

Even though mtDNA and chrY furnished us the initial portrait of human population structure and remain crucial to the studies revealing sex-biased migratory patterns, the single locus markers may ambiguously and too narrowly deduce the underlying past of the population divergences (Wakeley, 1996; Irwin, 2002; Rosenberg and Feldman, 2002). Indeed, they offer only two intriguing yet particular perspectives of the human population genetic history out of multiple plausible demographic histories of a population (Veeramah and Hammer, 2014). Henceforth, a collection of many supporting bits of evidence (from multiple loci) is needed to reconstruct the evolutionary past with a reasonable degree of broadness. Besides, the recombining quality of multi-locus markers makes them handy in see-through the multifaceted genetic changes brought upon by admixture and migration events. Technical advances, such as high-density genotyping arrays and parallel next-generation sequencing, triggered the current surge in genetic data from multiple genetic loci. Both provide massive new information by capturing the variations of a significantly higher number of independent/non-independent loci across the genome, which can be exploited with powerful statistical and analytical tools to deduce complex population demographic histories. The DNA hybridization and primer extension (microarray-based) high-density SNP genotyping generate a fraction of genetic data by sequencing a particular subsection of the genome. In comparison, massive parallel next-generation sequencing (NGS) facilitates the genetic data from the whole-genome sequence.

High-throughput SNP genotyping arrays were developed initially for genome-wide association studies (GWAS) of the biomedical community, aimed to map disease alleles by thoroughly searching for the genomic regions enriched in complex trait-related loci (Novembre and Ramachandran, 2011). The International HapMap project pioneered early genome-wide studies to show the genetic variation and linkage disequilibrium (LD) patterns in the human genome. HapMap phase I published data on 1 million SNPs by genotyping at least one common (minor allele frequency, MAF $\geq$ 0.05) SNP per 5 kb in each of 269 samples from four diverse populations across the globe representing Africa, Europe, and Asia (International Hap Map, 2005). The number of typed SNPs augmented to

3.1 million during HapMap Phase II when SNPs with MAF < 0.05 were genotyped on the same samples (The International HapMap Consortium et al., 2007), while Phase III added samples from seven more diverse populations (International HapMap Consortium, 2010). Later, the HUGO Pan-Asian SNP consortium added more SNP data of 1719 distinct individuals from 71 different populations of Asian origin (Consortium et al., 2009; Ngamphiw et al., 2011), substantially filling up the gap left by the HapMap project in population coverage from Asia. The variability of genome-wide SNPs has been successfully applied to reconstruct the genetic and demographic history of an ample number of populations from different corners of the globe, e.g., Europe (Novembre et al., 2008; Auton et al., 2009; Nelis et al., 2009; Kushniarevich et al., 2015; Tambets et al., 2018) and native Americans (Wang et al., 2007; Reich et al., 2012; Verdu et al., 2014; Raghavan et al., 2015). Several such studies also included underrepresented ethnic groups of South Asia (Reich et al., 2009; Metspalu et al., 2011; Moorjani et al., 2013; Basu et al., 2016), Africa and African-Americans (Tishkoff et al., 2009; Bryc et al., 2010), and Near and Middle East (Fernandes et al., 2019; Eaaswarkhanth et al., 2021; Mineta et al., 2021). Despite the advantages of SNP arrays in affordability and being the fastest to decode human population structure at higher resolution locally and globally, the intrinsic ascertainment bias arising mainly due to the design of the microarrays curbs the capability of genotyping arrays to infer evolutionary processes (Novembre and Ramachandran, 2011; Lachance and Tishkoff, 2013; Veeramah and Hammer, 2014). Generally, SNP arrays are used for genotyping a larger sample set using polymorphisms already discovered in small panels of individuals from a limited number of population groups. Consequently, they lack the genetic diversity extant in human populations worldwide (Lachance and Tishkoff, 2013). It happens due to the unequal distribution of genetic diversity across human population groups, and a small number of individuals allow capturing of variance in common alleles (with higher MAF) rather than rare ones (with low MAF) (Gravel et al., 2011). Typically, SNPs identified in discovery panels are not random but ones with higher MAF, i.e., common SNPs with intermediate occurrence (Novembre and Ramachandran, 2011). After the modern humans spread out of Africa, they have been exposed to successive bottlenecks and founder effects, resulting in lower genetic diversity among populations from Eurasia than Africa. Thus, it is highly possible to end up with flawed inferences when SNPs determined in one population are directed to genotype other populations (Rogers and Jorde, 1996; Eller, 2001). Enrichment of SNPs in some stretches of the genome, but their scarcity in other genomic regions of the genotyping array is another form of bias (Lachance and Tishkoff, 2013). Interestingly, ascertainment bias is sturdier in indels (insertion or deletion of nucleotides in the genome) but weaker in the case of microsatellites when compared to that of SNPs (Romero et al., 2009). Thus, the development of the next-generation sequencing (NGS) method, which has a great potential to get the better of most limitations presented by genotyping array, excited the scholars. NGS pools the contiguous DNA sequences from several or indeed all chromosomes and is supposed to be free of any ascertainment bias; nevertheless, various

quality control-related filtering procedures might still encourage a finer ascertainment bias (Novembre and Ramachandran, 2011; Veeramah and Hammer, 2014). The NGS leverage a significant upsurge in the efficiency to infer processes shaping the demography of human populations by exploiting either a full genome or a big chunk of the genome. The feature above, coupled with a downward trend in costs to get a high-quality full genome sequenced, is prompting more and more full genome-based studies unveiling complex demography of modern humans (Malaspinas et al., 2016; Mallick et al., 2016; Pagani et al., 2016; Chiang et al., 2018). Specifically, the 1000 genomes project (1KGP) launched in 2008 has added up almost 40 million novel variants (SNPs and indels), and almost 24% of these novel variants belong to South Asia, expanding the catalog of South Asian genetic variation tremendously (Auton et al., 2015). Phase I of 1KGP represented 1092 low coverage sequences and exome data mainly from sub-Saharan Africa, Europe, East Asia, and America (McVean et al., 2012). The sample set was later expanded up to around 1700 during phase II. Phase III of the 1KGP contributed genomic sequences of 2504 individuals that included additional samples from Africa and South Asia (Auton et al., 2015). Recently, the pilot phase of the ambitious project, the GenomeAsia100K Project, referred to as the GenomeAsia Pilot (GAsP) project, sequenced whole genomes of 1739 individuals from 219 populations across Asia to bridge the gap of underrepresentation of non-European individuals for genetic studies (Wall et al., 2019). Recognizing the importance of reference genome datasets required to characterize population-specific variation, the GAsP recruited samples across Asia. It included 724 individuals of South Asia thus, expanding the representation of diminished South Asian genetic variation beyond the sub-optimal level. However, the overall expenses, a cumulative sum of the cost related to the generation of high-quality genomes for a huge number of samples and the computational cost for several bioinformatics algorithms essential to analyze data efficiently (Mardis, 2011), is still a limiting factor to use NGS in regular studies based on large samples.

Nevertheless, the ever-growing multi-locus – autosomal and whole-genome – data of South Asia has rendered studying the population histories of the Indian subcontinent to an unprecedented level, identifying broader dimensions of the present-day genetic diversities of the region.

2.2. The power of aDNA to decipher the human demographic history

Availability of modern human genomes and different approaches analyzing them enabled a better understanding of human origin and evolution; for example, a study (Cann et al., 1987) postulated that modern humans originated in Africa. Subsequent methodological advances enabled the extraction and sequencing of aDNA from the relics of ancient humans (Rasmussen et al., 2010; Reich et al., 2012) and hominins (Green et al., 2010; Reich et al., 2010). Consequently, resolving the unsettled aspects of human history and evolution (Pääbo, 2015;

Llamas et al., 2017; Nielsen et al., 2017; Reich, 2018; Racimo et al., 2020). The evolving bioinformatics' and population genetic tools also enabled inferring the genetic components of ancient migrants and ancient admixture events that presumably shaped genetic variations and adaptations of the present-day human populations (Slatkin and Racimo, 2016; Dannemann and Racimo, 2018; Skoglund and Mathieson, 2018).

Nevertheless, obtaining reliable sequence data from aDNA used to be an uphill task until recently. The causal factors are post-mortem damage to DNA bases, lower quantity of endogenous genetic material, and smaller DNA fragments – a cumulative consequence of DNA degradation after the death of individuals (Veeramah and Hammer, 2014; Orlando et al., 2015). Contamination by several sources (including ancient and modern, humans and non-humans) is another vital concern that likely affects the power of interpretations made by downstream analyses of aDNA data (Kirsanow and Burger, 2012). However, technical improvements overcame caveats above more or less, enabling the preparation of aDNA sequencing libraries using double-stranded DNA (dsDNA) (Meyer and Kircher, 2010) or single-stranded DNA(ssDNA) (Gansauge and Meyer, 2013). Resulting aDNA libraries are then usually sequenced as it is via shotgun sequencing approach, or can alternatively be enriched for specific sequences via DNA capture – the enrichment of target aDNA sequences by hybridizing DNA of interest to the target probes (Briggs et al., 2009; Maricic et al., 2010; Carpenter et al., 2013; Fu et al., 2013). As a result, we see a surge in aDNA genome data not only across Eurasia but lately even from other parts of the globe with harsh climatic conditions for aDNA preservation, enhancing the perception of human demographic history to a deeper scale (Haak et al., 2015; Mathieson et al., 2015; Orlando et al., 2015; Fu et al., 2016; Lazaridis et al., 2016; Schlebusch et al., 2017; Järve et al., 2019; Narasimhan et al., 2019; Shinde et al., 2019; Lipson et al., 2020).

Unearthing aDNA from Africa has abetted to disentangle the distant past of human populations. aDNA studies (Llorente et al., 2015; Skoglund et al., 2017) discovered that mostly the Bantu speakers were responsible for the wide-spread farming in Africa, apart from finding the traces of Holocene interactions between Africa and Eurasia (Lazaridis et al., 2016; Skoglund et al., 2017; Fregel et al., 2018). Additional aDNA studies inferred that the deepest branch of modern humans lies in the southern African hunter-gatherers, who split from other populations ~ 250 KYA (Gronau et al., 2011; Mallick et al., 2016; Schlebusch et al., 2017). A recent study used aDNA from west African foragers, revealing ubiquitous admixture and three noticeable expansions; one such expansion mostly gave rise to no less than four key descents in the distant history of modern humans (Lipson et al., 2020). aDNA had significantly supported evidence of early divergence between ancestral West Eurasians and East Asians when the first ancient individual with a distinct genetic sharing to modern Europeans than East Asians was recorded (Seguin-Orlando et al., 2014). aDNA helped in reconstructing the complex history of Europeans. The AMH lived in Europe ~ 43 KYA, they had contributed a little, if any, to modern European genomes (Benazzi et al., 2011; Higham et al., 2011); a couple of recent studies also have added new insights

(Hajdinjak et al., 2021; Prüfer et al., 2021). The contemporary European genome is mostly a combination of three genetic components related to Mesolithic hunter-gatherers, Neolithic Anatolia, and Steppe peoples that arrived in Europe at different times (Lazaridis et al., 2014; Allentoft et al., 2015; Haak et al., 2015). The aDNA, furthermore, rendered the clue that farming in Europe was a consequence of the expansion of peoples from Anatolia (Gamba et al., 2014; Lazaridis et al., 2014, 2016; Sikora et al., 2014; Skoglund et al., 2014; Haak et al., 2015; Mathieson et al., 2015, 2018; Broushaki et al., 2016; Hofmanová et al., 2016; Lipson et al., 2017). East Asian aDNA revealed the ancient peoples of northeast China and southeast China (Fujian people) were distinct during the early Neolithic; however, the subsequent mixing during the late Neolithic spread of agriculture shaped the genetics of present-day East Asians who have predominant ancestry from the ancient northern group (Wang et al., 2020; Yang et al., 2020). Studies comprising ancient and modern human genomes have revealed four distinct ancestral components that arrived in the Americas in three assumed waves. The most recent wave was from Thule-related neo-Eskimos (~ 2 KYA), which followed the earlier Saqqaq-related Palaeo-Eskimo (~4.5 KYA), and the dispersal of Native Americans (before ~13 KYA) that diversified into two branches – the southern and northern branch (Reich et al., 2012; Raghavan et al., 2015; Moreno-Mayar et al., 2018; Scheib et al., 2018).

The aDNA specimens from South Pacific revealed that the Papuan ancestry observed in Remote Oceanians at present is plausibly the result of human interactions after the initial population settlement in far-off Oceania, which caused the spread of Papuans ancestry via the South Pacific (Skoglund et al., 2016).

Evolving aDNA technologies even allowed to sequence the first ancient individual from Indus Valley Civilization (IVC), a Bronze Age (BA) settlement of South Asia (Shinde et al., 2019). The study observed a lack of the Steppe ancestry (prevalent in northern India) in the BA Indus people and the ancient Iranian ancestry of IVC people deriving from a group that was ancestor to the Iranian farmers and hunter-gatherers. Another study (Narasimhan et al., 2019) suggested a plausible introduction of the Steppe ancestry into South Asia during the Middle to Late Bronze Age that also brought Indo-European language (Narasimhan et al., 2019). However, additional aDNA work with more and broader archaeological samples is now very much needed to substantiate further interesting conclusions drawn in these two studies.

2.2.1. aDNA revealing archaic admixture in modern humans

However, the most distinctively, aDNA sequencing is providing unprecedented discoveries through archaic genomes. For instance, comparative analysis of genetic material from both the modern humans and ancient archaic hominins delivered the revelation that ancestors of all contemporary non-Africans interbred with multiple archaic hominins, including Neanderthals and Denisovans (Green et al., 2010; Reich et al., 2010; Meyer et al., 2012; Fu et al., 2014; Teixeira and

Cooper, 2019). The first successfully sequenced archaic genome was that of Neanderthals, suggesting interbreeding between the ancestors of all contemporary non-Africans and Neanderthals at ~50–60 KYA (Green et al., 2010; Sankararaman et al., 2012; Fu et al., 2014). Follow-up studies identified that ~2–4% genome of all modern non-African humans come from Neanderthal genomes, plausibly a result of introgression from European Neanderthals (Croatian Vindija cave) rather than the Altaian one (Prüfer et al., 2014, 2017). Lately, a sequence of ~ 60–80 KYA old Neanderthal female from Chagyrskaya that lived near the Denisova cave of Altai mountains (Mafessoni et al., 2020) revealed its higher genetic intimacy to the Vindija cave Neanderthal than their Siberian counterparts, suggesting a likely replacement of the Siberian Neanderthals by the Western ones. Whereas the genome sequence of Denisovans – archaic hominin sister group of Neanderthals – revealed the interbreeding of Denisovans with ancestral people of Sahul around 45–49 KYA (Malaspinas et al., 2016). Different studies observed ~3–6% genetic contribution of Denisovan genome to the present-day Papuan New Guineans and Australian aboriginals, in addition to a lower yet traceable ancestry to the East Asians and South Asians (Reich et al., 2010; Skoglund and Jakobsson, 2011; Meyer et al., 2012; Qin and Stoneking, 2015; Malaspinas et al., 2016; Sankararaman et al., 2016; Browning et al., 2018; Jacobs et al., 2019). Surprisingly, a recent massive study of more than 27,000 genomes of Icelanders revealed the presence of Denisovan genome fragments despite this small northern Atlantic island population indeed very far away from Oceania (Skov et al., 2020). However, the study remains undecided whether the attributed Denisovan fragments in Icelanders' DNA testifies about the direct admixture of the remote ancestors of Icelanders with Denisovans or were carried by the Neanderthals that plausibly derived from earlier Neanderthal-Denisovan admixture events.

Nevertheless, several studies anticipated additional interbreeding events; for example, a plausible genetic contribution from eastern Neanderthals and morphologically unknown-archaic group to Denisovans (Prüfer et al., 2014), and from early modern humans to Neanderthals (Kuhlwilm et al., 2016; Posth et al., 2017). Studies also found now (a weak) signal of Neanderthal fragments even in Africans that might well be the consequence of well-known backflow of modern humans who admixed with Neanderthals in Eurasia and carried this signal to Africa (Hsieh et al., 2016; Xu et al., 2017; Durvasula and Sankararaman, 2019; Chen et al., 2020).

Recent studies have also supported an intricate pattern of archaic introgression in Asia (Mondal et al., 2019; Teixeira and Cooper, 2019). That, considering the region's evidence of occupation by numerous hominin groups (Détroit et al., 2019), suggests likely interbreeding with unknown archaic hominins apart from the known Neanderthal and Denisovan ones.

2.3. Methods to detect genome-wide patterns of population structure and demographic history

Genetic distances between modern populations are majorly shaped by genetic drift, which is commonly stronger in small populations and, in the long term, creates pronounced differences in the presence of natural or geographic barriers. The differences among populations involve only a tiny percentage of the existing genetic variation, with the most variation being within them (Lewontin, 1972). An ever-growing body of genome-wide data across the globe and advances in genetic analysis methods have empowered the studies to shed light on evolutionary processes that shaped the demographic history of modern humans. Here, I survey different approaches employed to analyze such genomic data to trace population structure and its dynamics, identify and date admixture, and shared haplotype-based models describing some aspects of human demographic history.

2.3.1. Mean pairwise genetic distance (F_{ST})

Statistics measuring genetic distance among populations provide an insight into the proportional relationship of populations – inferring greater genetic distance correlates to a greater evolutionary distance between populations. F_{ST} or mean pairwise distance (Weir and Cockerham, 1984; Holsinger and Weir, 2009) is a statistic that measures the amount of genetic differentiation among a set of populations. F_{ST} utilizes the uneven distribution of allele frequencies among populations to compare the amount of extant genetic variation within pairs of subpopulations corresponding to the entire population. The value of F_{ST} can vary between 0 to 1, depending on how genetically differentiated populations are, with higher values indicating higher differentiation.

2.3.2. Spatial techniques to infer ancestry at a global level

Principal Component Analysis (PCA) is an algebraic tool that, as applied to genotype data, assigns an individual's data onto a small set of dimensions. These dimensions are characterized by maximal separation of the data reducing high dimensional covariance matrix data to independent principal components (PCs) or eigenvectors that comprise the majority of variation present in the complete data. Generally, PCA is applied to many individuals' genomic data, but only a few top PCs are considered for visualization **(Figure 1a)**. However, given the high similarity characterizing our species, the number of variables, and the degree of complexity of the underlying genetic structure, the variance explained by these top components is meager (Novembre and Stephens, 2008). When a PCA is plotted with the first few PCs, it displays clustering between genetically close individuals and more separation between individuals that are genetically less similar, where a group at an intermediate location might not necessarily be the result of genetic

admixture (Novembre and Stephens, 2008; McVean, 2009; Schraiber and Akey, 2015; Wangkumhang and Hellenthal, 2018). In general, PCA may capture the geographical pattern of migrations among populations, underlying the correlations between PCs and geographical axes (Cavalli-Sforza et al., 1994; Reich et al., 2008).

Nevertheless, sample size can affect PCA, and care must be taken while interpreting PCA results since the gradient of PCs might result from mathematical artifacts or recurrent bottlenecks caused by range expansions (Novembre and Stephens, 2008; François et al., 2010; Klopfstein et al., 2006; McVean, 2009). A few methods have been developed recently to address caveats in PCA interpretations, exploiting the isolation by distance model (i.e., genetic similarities among populations decrease with increasing geographic distance). SpaceMix (Bradburd et al., 2016) is one such statistical approach to building a geogenetic map embedding populations based on their genetic closeness rather than geographical proximity. Thereby, the approach represents a pattern of long-term gene flow among the populations, spotting groups with more highly correlated allelic frequencies than expected under the isolation-by-distance model (Bradburd et al., 2016; Novembre and Peter, 2016; Wangkumhang and Hellenthal, 2018). The Estimated Effective Migration Surface or EEMS (Petkova et al., 2016) approach, on the other hand, predicts the discrepancy in the effective migration across a surface by relating the genetic variations among individuals to the geographical distances of their sampling locations. The method assumes a stepping stone model for migration between adjoining population groups (Novembre and Peter, 2016; Petkova et al., 2016; Wangkumhang and Hellenthal, 2018). Therefore, both approaches can be helpful to infer unexpected genetic correlations among populations that arise either due to the admixture from a group farther apart or due to migration barriers between populations.

2.3.3. Clustering algorithms-based methods

Although PCA-based approaches are a ubiquitous way to visualize genetic structure in genomic data sets, it is hard to retrieve estimates of global ancestries for tested individuals from such statistical tools. Therefore, several clustering (structure-like) methods of population assignments have been developed. These methods search for structure by grouping individuals with shared frequencies of common alleles, mutually in Hardy Weinberg equilibrium (HWE) (Schraiber and Akey, 2015). Different clustering algorithms are analogous in their aim to group individuals into various distinct populations or clusters (K), where each cluster represents individuals sharing a typical genetic pattern (or allele frequencies). The Bayesian-based approach, STRUCTURE (Pritchard et al., 2000), applies Markov chain Monte Carlo (MCMC) on allele frequencies of populations allocating a single individual to a cluster (or population) after the calculated percentage of ancestry from each discrete population is averaged over the entire genome of the individual. The method might be arbitrarily sluggish for larger data sets. Therefore, to analyze more extensive genomic data sets efficiently, faster tools like

ADMIXTURE (Alexander et al., 2009) and FRAPPE (Tang et al., 2005) were developed using maximum likelihood methods rather than MCMC. Whereas, fastSTRUCTURE (Raj et al., 2014) uses a variational approximation approach of Bayesian that provides a fast and deterministic substitute to Monte Carlo methods. All these techniques undertake unlinked or independent loci and serve as global ancestry estimation methods. These algorithms generally require a pre-defined number of anticipated populations or clusters K and subsequently assign a fraction of each individual's genome to the suggested number of clusters K **(Figure 1b)**. However, choosing the optimal number of source populations is challenging. In this context, ADMIXTURE implements a cross-validation method to select the number of populations (Alexander et al., 2009). On the contrary, fastSTRUCTURE uses heuristic scores to find sufficient population numbers (Raj et al., 2014). It is also plausible to over-interpret STRUCTURE/ADMIXTURE results as many different demographic settings might end up mainly yielding similar cluster distribution (Pritchard et al., 2000; Novembre, 2016; Lawson et al., 2018; Wangkumhang and Hellenthal, 2018). Therefore, a mixed pattern of clusters does not always represent actual admixture, e.g., individuals with a history of admixture with an unknown ghost population. Moreover, assigning a group of individuals to a single cluster does not indicate that individuals have not undergone admixture, e.g., if a sister population (Kalash) is way more privately drifted due to a recent substantial population bottleneck (Ayub et al., 2015a). Consequently, STRUCTURE or ADMIXTURE results must be corroborated by other methods that follow varied modeling assumptions such as TreeMix (Pickrell and Pritchard, 2012), fineSTRUCTURE (Lawson et al., 2012), f_3 and D statistics (Patterson et al., 2012), etc., to conclude patterns of population mixing and demographic histories decisively.

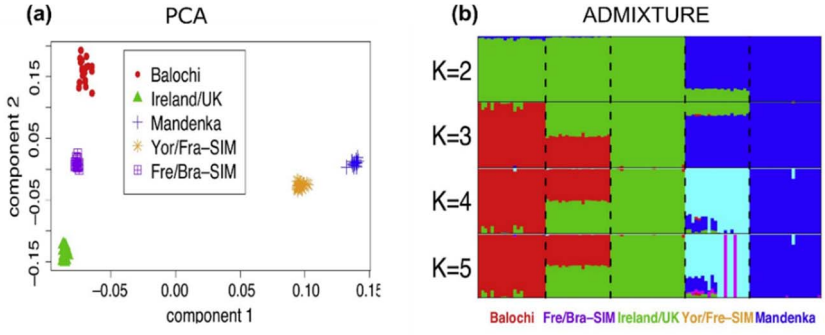

Figure 1. PCA and STRUCTURE like analyses in simulated data.

(a) Principal component analysis (PCA) via EIGENSRTAT where each dot represents an individual, **(b)** ADMIXTURE for cluster numbers K = 2–5 (columns = individuals, colors = clusters). Adapted from Figure 1 in Wangkumhang and Hellenthal 2018 published in *Curr Opin in Genet & Develop,* with permission from "Elsevier".

2.3.4. Inferring a fine-grained picture of admixture history from haplotypes

In general, genomes of admixed individuals result from the blending of several DNA segments inherited from individuals of diverse origins, that keep dividing into smaller chunks with time due to recombination. Original admixing sources for inherited DNA segments can be identified by comparing the genome of the admixed individual to that of a given set of sources via chromosome painting or identical-by-descent (IBD) tracts (Novembre and Stephens, 2008; McVean, 2009; Schraiber and Akey, 2015; Wangkumhang and Hellenthal, 2018). Later, the sizes and length distribution of these DNA tracts are compared to the ones expected under different migration models, and the date of admixture event and ancestry proportions are inferred, outlining the admixture history at a more fine-grained level (Gravel, 2012; Churchhouse and Marchini, 2013; Novembre and Stephens, 2008; McVean, 2009; Schraiber and Akey, 2015; Wangkumhang and Hellenthal, 2018).

2.3.4.1. Approaches of local ancestry inference

Many different local ancestry (the genetic ancestry of an individual at a genomic location) methods have been developed to model the ancestry distribution of DNA tracts or haplotypes (Tang et al., 2006; Sankararaman et al., 2008; Price et al., 2009; Baran et al., 2012; Omberg et al., 2012; Churchhouse and Marchini, 2013; Maples et al., 2013). These methods vary depending on whether they need pre-phased data, the number of source populations allowed, and if they could model the loci in high LD (i.e., closely associated haplotypes). Most of the methods capture the ancestral haplotype structure within the generative framework of Li and Stephen's haplotype-copying model (Li and Stephens, 2003). Li and Stephen's model makes the fast approximation of the coalescence of DNA segments considering recombination events and estimates the target genome as a dense mixture of predefined source genomes, applying the hidden Markov model (HMM) (Schraiber and Akey, 2015). Such approaches can directly model LD for tightly linked markers, therefore, are capable of distinguishing genetically close populations (Baran et al., 2012), *e.g.,* HAPMIX (Price et al., 2009) and HAPAA (Sundquist et al., 2008). However, HAPMIX and HAPAA have limitations in that they allow only two predefined source populations at a time and are computationally intensive.

The presumption of a parametric population genetic model is a caveat for the tools above, unlike several non-parametric local ancestry methods. Non-parametric local ancestry methods do not follow Li and Stephen's generalized framework and work on a window-based approach, although they, too, need a set of reference populations (Schraiber and Akey, 2015). These methods are fast, allow several source populations, and are capable of unraveling exceptionally complex demographies (Churchhouse and Marchini, 2013; Schraiber and Akey, 2015).

Such methods, e.g., LAMP (Sankararaman et al., 2008), PCAdmix (Brisbin et al., 2012), Saber (Tang et al., 2006), SupportMix (Omberg et al., 2012), WINPOP (Paşaniuc et al., 2009) vary in their approaches. Nevertheless, these approaches mostly do not model background LD and break the individual genomes into windows, grouping them according to each window's assignment to a reference panel. However, the window-based approaches LAMP-LD (Baran et al., 2012) and MultiMix (Churchhouse and Marchini, 2013) can also model background LD. Nonetheless, we need to be very careful in choosing the window size for such methods as it can significantly affect the results.

Most local ancestry algorithms follow a generative approach that attempts to find, indirectly, the hidden parameters from a joint distribution of data. It requires a set of reference populations first to estimate the shared reliance of observed (alleles) and unobserved (ancestries) variables in data and then assesses the dependency of unobserved variables on observed variables using Bayes' rule via HMM (Maples et al., 2013; Padhukasahasram, 2014). Thus, restricting these methods from exploiting the inherent ancestry of admixed samples. To overcome these issues, a discriminative-approach (that finds boundary to distinguish the data into different classes) based method RFMix (Maples et al., 2013) was developed. RFMix models the probabilistic dependence of unobserved variables (ancestry) directly on observed variables (alleles) and estimates the ancestry of an admixed chromosome. Discriminative approaches are supposedly more accurate and faster than the generative approaches when the data set is large (Ng and Jordan, 2001). However, regardless of the approaches, local ancestry methods have been found applicable to studies related to map disease genes via admixture mapping (Hoggart et al., 2004; Reich et al., 2005), pharmacogenomics (Yang et al., 2011), selection (Tang et al., 2007; Jin et al., 2012), and demographic inferences (Bryc et al., 2010; Kidd et al., 2012).

2.3.4.2. Approaches that side-step local ancestry inference

Determining local ancestries when ancestral populations (sources) are poorly differentiated remains a challenge. Thus, alternative approaches (e.g., fineSTRUCTURE, GLOBETROTTER) that can identify admixture events without requiring source proxies to be defined are helpful (Lawson et al., 2012; Hellenthal et al., 2014; Busby et al., 2015; Leslie et al., 2015). fineSTRUCTURE-ChromoPainter framework (Lawson et al., 2012) uses Li and Stephen's haplotype copying model (Li and Stephens, 2003) to detect shared haplotypes, revealing fine-scale ancestral relationships among individuals. The method can infer more than two sources and may characterize population relationships at fine scales. fineSTRUCTURE sums up the output of ChromoPainter (Lawson et al., 2012) that reconstructs chromosomes of each individual as a combination of copying chunks received from all other individuals in a co-ancestry matrix **(Figure 2)**. This co-ancestry matrix can be used for PCA to identify more discrete clustering of related individuals and reconstruct the tree-like hierarchical relationships among different

population clusters using fineSTRUCTURE (Lawson et al., 2012; Busby et al., 2015). fineSTRUCTURE can also infer distinct and hidden signals of population structure by applying the mixture modeling of copying profiles on the detected shared haplotypes using GLOBETROTTER (Hellenthal et al., 2014), and NNLS approaches (Leslie et al., 2015; Montinaro et al., 2015).

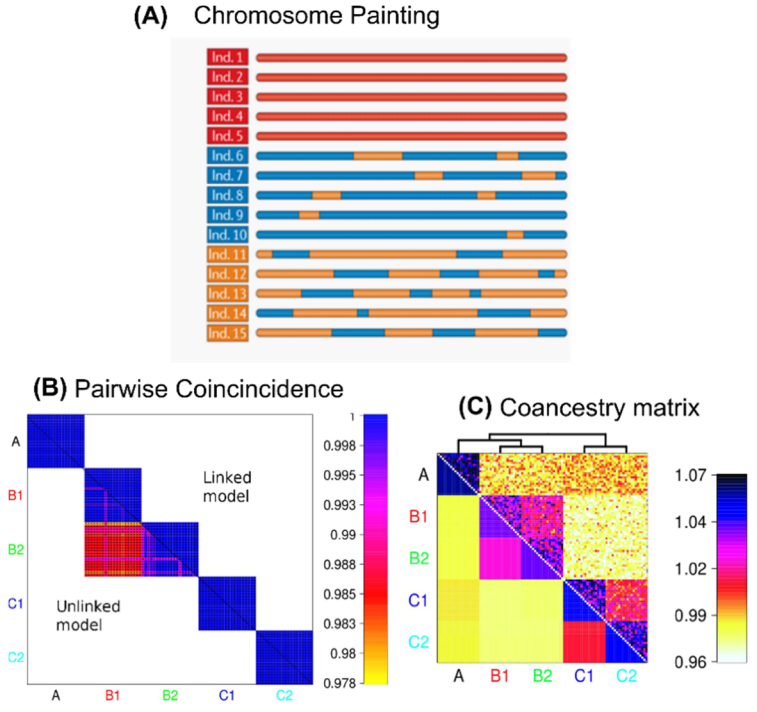

Figure 2. Inferring fine-scale ancestry from simulated data.

(a) Chromosomal painting shows the specific tracts of sequences inherited from ancestors in each population, **(b)** Pairwise coincidence matrix output by fineSTRUCTURE using chunk counts calculated using (top right) the linked and (bottom left) unlinked model, for the datasets from Lawson et al. 2012. The coloring represents the posterior coincidence probability (which does not drop below 97%) and the dots represent the maximum a posteriori (MAP) probability state, (c) Aggregated coancestry matrix (bottom left, normalized to have row mean 1) for the linked model dataset (top right) from Lawson et al. 2012 shown with the inferred MAP tree (top). Adapted with permission through Copyright Clearance Centre, Inc. from "Springer Nature" with minor alterations in Figure 2e from Schraiber and Akey (2015) published in *Nat Rev Genet,* and from Figure 3 of Lawson et al. (2012) published in *PLoS Genet* under terms of Creative Commons Attribution License.

GLOBETROTTER (Hellenthal et al., 2014) is an extension of the fineSTRUCTURE algorithm to account for the genetic heritage from groups that are not available in the data. GLOBETROTTER reconstructs each admixture source as a combination of DNA segments from all donor individuals, therefore, inferring the specific features of admixture (Busby et al., 2015; Leslie et al., 2015; Montinaro et al., 2015; Schraiber and Akey, 2015; Wangkumhang and Hellenthal, 2018). While in the NNLS approach, a modified version of the non-negative least square (NNLS) method is applied on the copying vectors obtained from ChromoPainter to reconstruct the ancestry profile of each cluster or individual (Leslie et al., 2015; Montinaro et al., 2015).

2.3.4.3. Identical-by-descent (IBD) approach

IBD is another approach that characterizes fine-scale population structure appropriately using extended regions of identical sequences (haplotypes) shared between pairs of genomes (Palamara et al., 2012). When a genomic region shared between two or more individuals contains no mutation that distinguishes them, it is called identical-by-state (IBS). If this identical region was inherited from a common ancestor without recombination, it is known as IBD. All individuals of a fixed population are related to a common ancestor going back in time, consequently contain shared DNA segments that are IBD due to their inheritance from a common ancestor assuming no breakage by recombination; thus, modeling of recombination is mostly avoided giving it more power (Schraiber and Akey, 2015). However, during meiosis, recombination breaks these IBD segments into smaller segments that decay exponentially with the number of past generations from their most recent common ancestor (MRCA). Henceforth, the length of IBD blocks can infer the time to MRCA, with longer IBD blocks associated with a very recent common ancestor, and shorter IBD blocks represent the coancestry due to ancient origin (Ringbauer et al., 2017). Establishing the shared segments as IBD segments is a challenging task due to possible errors in assuming a segment as IBD when created by nucleotides that are identical by state; which might happen when a region is conserved or in high LD, thus detecting false positive IBD segments (Browning and Browning, 2010; Schraiber and Akey, 2015). Nevertheless, technical advancements have enabled the detection of a range of IBD tracts between pairs of genomes, evading the issue of detecting false positives and small segments related to ancient origin, by recommending to use the IBD blocks no shorter than a certain map length (Browning and Browning, 2013; Schraiber and Akey, 2015). Methods such as Beagle IBD (Browning and Browning, 2010) and RELATE (Albrechtsen et al., 2009) use a probabilistic approach to determine IBD via HMM and are computationally extensive. Whereas, methods to detect IBD in large data sets follow a non-probabilistic approach, with an efficiency that increases proportionately with sample sizes, include fast IBD (Browning and Browning, 2011), Refined IBD (Browning and Browning, 2013), and GERMLINE (Gusev et al., 2009). Refined IBD has higher accuracy due to its advantages over the other two methods because; it uses the data more efficiently

by integrating LD modeling (Browning and Browning, 2013). Additionally, Refined IBD's probabilistic approach-based fine-tuning can better deal with ambiguity.

Long IBD blocks can be harnessed to describe a range of recent demographic features, e.g., to extract information on recent population sizes by fitting different migration models with IBD (Palamara et al., 2012; Browning and Browning, 2015), and to infer the historical migration rates of populations (Palamara and Pe'er, 2013). Nevertheless, studies have also utilized shorter IBD tracts to infer demographic events of deeper timescales (Li and Durbin, 2011; Harris and Nielsen, 2013). Recently, a study developed a novel probabilistic approach based on IBD sharing, detecting the introgressed sequences of Neanderthals into the modern Africans (Chen et al., 2020). Another study applied IBD and local ancestry approaches to get insight into the overall and ancestry-specific *Ne* during the past hundred generations for several admixed American populations (Browning et al., 2018a).

2.3.5. Dating admixture using pattern of LD decay

Most of the population genetic methods either use allele frequency variation among individuals (e.g., PCA, ADMIXTURE, *f*-stats) or directly infer local ancestry segments in chromosomes (e.g., PCAdmix, HAPMIX) to identify and reconstruct historical admixture events that occurred between population groups. However, inferring the date of the admixture event is tough from these methods, as PCA and STRUCTURE-like analyses assume admixed populations are genetic intermediates of plausible source groups (Loh et al., 2013). While *f*- and *D*- stats are unable to infer the timing of an admixture event due to the dependency of genetic drift on population size (Loh et al., 2013). Although local ancestry methods might infer the admixture date by exploiting the distribution of DNA tracts, minuscule tracts due to consecutive recombinations or in case populations in the reference set are more diverged than original mixing sources, might undermine the possibility to date the admixture event (Loh et al., 2013; Wangkumhang and Hellenthal, 2018). To solve the issue, researchers develop methods that apply an alternative approach of using the extent of exponential decay in linkage disequilibrium among loci (i.e., Admixture LD) (Moorjani et al., 2011; Loh et al., 2013). When a population inherits a chromosomal chunk with correlated loci from one of the source groups during the past admixture events, the association among loci might be retained even after several recent mixings, resulting in admixture LD (Chakraborty and Weiss, 1988). Both ROLLOFF (Moorjani et al., 2011; Patterson et al., 2012) and ALDER (Loh et al., 2013) compute the decay in admixture LD vis-à-vis to the genetic distance between a pair of SNP markers, weighted by the variation in allele frequency between source groups (Moorjani et al., 2011; Loh et al., 2013; Wangkumhang and Hellenthal, 2018). Since recombination breaks down the admixture induced associations among loci, the weighted LD between two markers in an admixed population decays exponentially (i.e., weighted LD curve) at a constant rate, providing the age of the admixture event as implemented in ROLLOFF (Moorjani et al., 2011; Patterson et al., 2012).

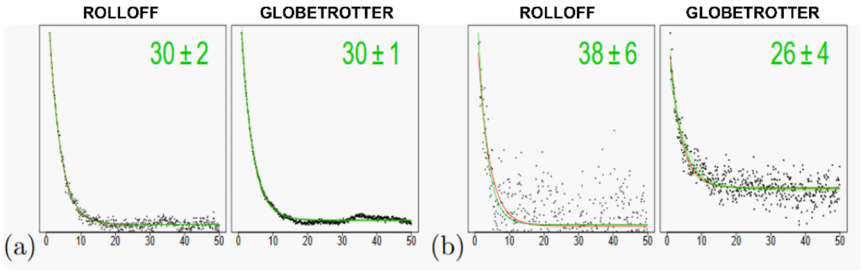

Figure 3. Dating admixture applying ROLLOFF and GLOBETROTTER to simulated data.

Coancestry curves for ROLLOFF (left) versus GLOBETROTTER (right) for the following simulated admixture events: **(a)** 80% contribution from Brahui plus 20% contribution from Yoruba,30 generations ago; **(b)** 50% French + 50% Brahui, 30 gen ago. Inferred dates and standard errors are given at the top right in each plot. Green lines give exponential decay curves with rates equal to inferred dates of admixture; red lines give exponential decay curves with rates equal to actual admixture dates – the coancestry curve for the surrogate group representing the minority admixing source. ROLLOFF standard errors were calculated using the Weighted Block Jackknife procedure, while the GLOBETROTTER method took the standard deviation across 100 bootstrap re-samples. Adapted with minor modification from Figure S8 in Hellenthal et al. (2014) published in *Science*, with permission from "AAAS".

However, the requirement to pre-specify the donor for each admixing source in ROLLOFF and ALDER is an issue when the source populations intermixed to give the target admixed population are unknown or have no proper representative among modern groups. Therefore, the haplotype-based approach GLOBETROTTER is advantageous, which assumes the admixing sources as a genetic mix of all donor groups; thus, pre-defined donor per source might not be used in specific scenarios (Hellenthal et al., 2014; Wangkumhang and Hellenthal, 2018). The approach of estimating the association between haplotype segments gives GLOBETROTTER extra power **(Figure 3)**, though it comes with a cost of computational time. This haplotype-based method can detect if more than two sources participated in the same admixture event, in addition to identifying multiple admixture events (Hellenthal et al., 2014; Wangkumhang and Hellenthal, 2018). However, the upper limit of admixture events that GLOBETROTTER can detect is lower (~150 generations) than for ALDER (~250 generations) and ROLLOFF (~300 generations). In general, most of the dating approaches assume the admixed target population has undergone one massive pulse of mixing, followed by random mating among individuals of the admixed group. However, if the admixed population has a history of constant mixing at different time intervals, these methods might detect only the most recent admixture event (Loh et al., 2013; Hellenthal et al., 2014; Wangkumhang and Hellenthal, 2018).

2.3.6. Identifying admixture using drift parameters

Patterns of common genetic variation between groups of populations can be informative of admixture and human populations history (Pickrell and Reich, 2014; Schraiber and Akey, 2015). One approach to analyze the shared history of populations is *f*-statistics (Patterson et al., 2012), pioneered by Reich et al. (Reich et al., 2009) to establish the origin of highly structured Indian populations. Due to its ability to study the widespread admixture events of human population history (Novembre and Peter, 2016), *f*-statistics has become a prominent tool for modern as well as ancient DNA studies (Green et al., 2010; Metspalu et al., 2011; Rasmussen et al., 2011; Reich et al., 2012; Allentoft et al., 2015; Haak et al., 2015; Lazaridis et al., 2016). *f*-statistics is a methodological framework for fitting the admixture events to the frequencies of genetic variants observed in several populations. Genetic drift shared between sets of two, three, or four populations can be estimated as f_2, f_3, and f_4, under the assumption that populations that share a similar level of genetic drift conceivably have a common evolutionary history (Patterson et al., 2012; Peter, 2016; Harris and DeGiorgio, 2017).

The f_2-statistic estimates the differentiation between two populations like Wright's F_{ST}, a long-term practice to explain the population relationships. The f_2 value refers to the branch length between two populations, applied to fit the edges (i.e., one population is a successor of another) in an admixture graph (Patterson et al., 2012; Peter, 2016; Harris and DeGiorgio, 2017).

The f_3-statistic or f_3 admixture test is a three-population test to analyze the genetic differentiation of a target population (A) from two other populations, B and C, to test if population A is derived from the admixture event between B and C **(Figure 4A)**. If population A is the result of an admixture event between two populations B and C, the population A's allele frequency (pA) of an SNP marker at a locus should be intermediate to the allelic frequencies (pB, pC) of that marker in populations B and C (Patterson et al., 2012; Peter, 2016; Harris and DeGiorgio, 2017; Wangkumhang and Hellenthal, 2018). Therefore, f_3 (A; B, C) is defined as the product of frequency differences between population A and B, and between population A and C, i.e., f_3 (A; B, C) = (pA-pB) (pA-pC), normalized and averaged over all SNPs. Since genetic drift between two populations is defined by the frequency difference between them, f_3 (A; B, C) is proportional to the relative genetic drift (or branch length) between populations A and B, and A and C. When population A is not a descendant of the admixture between groups (similar to) B and C, the genetic drift along the lineage that gave rise to population A after divergence from the ancestor group must be more than 0, i.e., the value of f_3 (A; B, C) > 0. If population A descends from admixture between population B and C, its allele frequency will fall between pB and pC, resulting in a significantly negative value for f_3 (A; B, C). f_3 (A; B, C) significance is evaluated by applying a jack-knife resampling method of independent chromosomal blocks. Nevertheless, caution should be taken when a population is privately drifted (due to genetic isolation) (Ayub et al., 2015), thus masking the admixture signal, i.e., a non-negative f_3 value (Patterson et al., 2012; Reich et al., 2012;

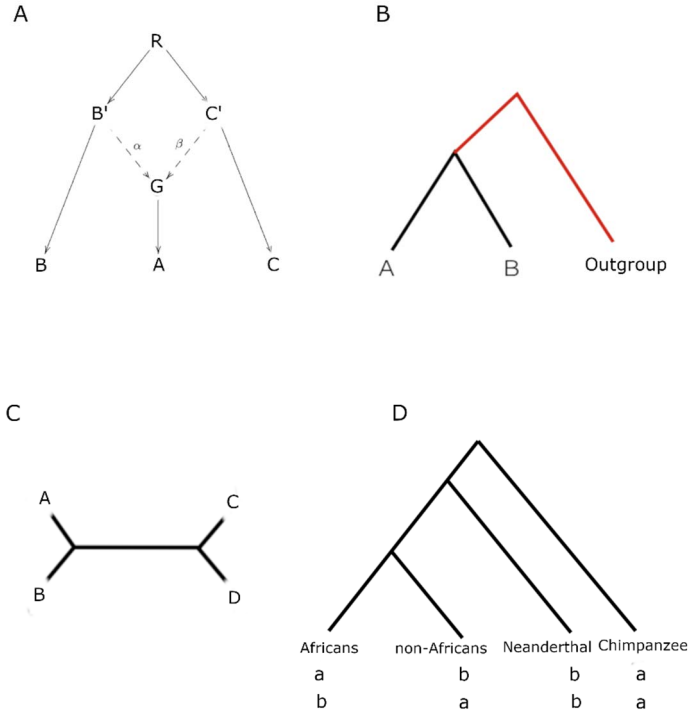

Figure 4. Representation of f- and D-statistics.

(A) Phylogenetic tree of simple admixture where B' and C' split from a root population R. Population A's ancestral lineage G was formed by admixture (between B' and C') in proportions α: β, (β = 1–α). Contemporary populations A, B, C are formed by drift from their ancestors (G, B' and C' respectively), thus f_3(A; B, C) will give negative value; **(B)** Outgroup statistic f_3 (Outgroup; A, B) measures the closeness between populations A and B, as measured from the outgroup, the red color indicates the branch length from outgroup to the common ancestor of A and B; **(C)** Unrooted four-population tree topology showing the relationships between four populations wherein A, B and C, D are forming two individual clades. The result f_4 (A, B; C, D) = 0 indicate that the tree is consistent, while a value beyond 0 indicates otherwise; **(D)** *D*-statistic representing a rooted four-population asymmetric tree relationship between modern humans (Africans, non-Africans) and Archaic hominins (Neanderthal) using Chimpanzee as the outgroup, the same can be applied for any four population combinations. abba configuration shows derived allele sharing between non-Africans and Neanderthal, and baba configuration indicates derived allele sharing between Africans and Neanderthal. Recreated based on the "Genetics Society of America" publisher's Patterson et al. 2012 in *Genetics,* and "AAAS" publisher's Green et al. 2010 in *Science*.

Wangkumhang and Hellenthal, 2018; Pathak, 2020). A form of the f_3-statistic, f_3 (Outgroup; A, B), referred to as outgroup-f_3 **(Figure 4B),** was developed by Raghavan et al. (Raghavan et al., 2014) to find the genetic closeness between a particular ancient group and each population from a referred list of modern populations. Generally, the Yoruba or Mbuti-Pygmies are used as an outgroup proxy for all non-African modern humans, assuming that the genetic drift in the lineage related to the outgroup is constant for all non-Africans (e.g., A and B) until there is some gene flow between A and B (shared genetic history). Thus, a higher f_3 (Outgroup; A, B) value infers relatively more common genetic history between populations A and B. Despite being a distance-based method like Wright's F_{ST}, the outgroup-f_3 is less affected by the genetic drift in the test population.

Another method to validate a proposed tree topology by measuring the shared drift among four populations is a test of treeness or the f_4-statistic. Generally, in an unrooted tree comprising four populations (A, B, C, D) with allele frequencies pA, pB, pC, and pD, respectively, the f_4-statistic formally evaluate whether populations (A, B) and (C, D) fit in two separate clades **(Figure 4C)**. Indeed, the f_4-statistic use the allele frequency differences between all four populations (A-D) (Reich et al., 2009) by calculating (pA-pB) and (pC-pD), averaged across all loci, and validate if populations (A, B) and (C, D) form clades or not.

In case of correct ((A,B), (C,D)) topology, the f_4-statistic should be zero because the allelic frequency difference (pA-pB) is not affected by (pC-pD). A significantly positive statistic f_4 (A,B; C,D) suggests a higher genetic relatedness between populations A and C (or B and D), whereas a significantly negative f_4 value infers more ancestral relation between B and C (or A and D). The resulting incorrect tree topology indicates significant shared ancestry between clades due to a plausible admixture scenario. Thus, the f_4-statistic is an efficient tool to test admixture, but with the inability to specify the direction of mixing. Another caveat of statistic f_4 is that it might infer, wrongly, that admixture did not occur when population D had received an equal amount of ancestry from both populations B and C (considering population A is an outgroup); while the f_3-statistic remains unaffected and suggests admixture did happen (Patterson et al., 2012; Reich et al., 2012; Harris and DeGiorgio, 2017; Pathak, 2020). The admixture proportion (p) of a test population (A), derived from the source population (C), can be estimated by f_4 ratio by dividing two f_4-statistic values generated from five sources in a form $p = f_4$(B,O; A,D)/ f_4(B,O; C,D) (Reich et al., 2009, 2012; Patterson et al., 2012; Harris and DeGiorgio, 2017).

The *D*-statistic is another four-population test that identifies the interbreeding between three in-group populations, based on an outgroup-rooted tree of four populations. *D*-statistic was initially presented (Green et al., 2010) to detect Neanderthal introgression in modern humans using sequencing data. Whereas extended *D*-statistic (Durand et al., 2011; Patterson et al., 2012) can be applied on genotyping data to test admixture even in the absence of archaic specimens; thus, it can be applied to any combination of four populations. The *D*-statistic is also known as the abba/baba approach, where $D = N_{baba} - N_{abba} / N_{baba} + N_{abba}$. N_{baba} represents the number of allelic sites shared between Africans and Neanderthal

(Figure 4D), while N_{abba} represents allelic sites shared between non-Africans and Neanderthal (Patterson et al., 2012; Harris and DeGiorgio, 2017). A significant positive value of the *D*-statistic indicates excess allele sharing (or gene flow) between Africans and Neanderthal. A significantly negative value suggests non-Africans and Neanderthal share higher alleles. Unlike the f_4-statistic, the value of the *D*-statistic lies between –1 and 1 due to the underlying normalization process; nonetheless, both *D* and f_4-statistics are robust to ascertainment bias (Patterson et al., 2012). However, the significance of the *D* value for a correct phylogeny is conditional to the *Z*-score > 3 or < –3 obtained via a weighted-block jackknife approach, like for *f*-statistics.

Subsequently, additional methods, *qpWave,* and *qpAdm* were developed applying common ideas linked to the f_4 statistic (Reich et al., 2012; Moorjani et al., 2013; Lazaridis et al., 2014; Haak et al., 2015). *qpWave* identifies the minimum number of reference populations possibly involved in the admixture history of ancestors of a specific test population. At the same time, *qpAdm* computes the ancestry proportions contributed by these sources to the test population. *qpAdm* framework requires a set of test and source populations – referred to as 'left' populations in addition to a set of 'right' or reference populations that are differentially related to populations in the 'left' list and calculates a matrix of *f*-statistic. If the source populations derive from n different ancestral groups, the matrix of the *f*-statistics will have a rank n-1 (Haak et al., 2015). An admixture model is deemed credible if it is not rejected ($p > 0.05$), i.e., adding the test population to the 'left' list does not require additional ancestral sources; thus, matrix rank remains unchanged. The calculated ancestry proportions should also fall within the range of 0–1 (Haak et al., 2015; Harney et al., 2020). Though *qpAdm* is robust to low sample size, missing data, and ascertainment bias; plausible rejection of true admixture models when many populations are analyzed, and misleading interpretation in case of continuous gene flow raise concerns (Harney et al., 2020). Nevertheless, *qpAdm* has been applied in several ancient DNA studies explaining the ancient ancestral components for different populations living today (Reich et al., 2012; Haak et al., 2015; Lazaridis et al., 2016; Damgaard et al., 2018; Narasimhan et al., 2019).

2.3.7. Tree-based tools of population splits and admixture

Different graph construction tools that infer tree-based topologies by relating several populations identifying population splits have also been developed. These methods (e.g., *qpGraph*, *MixMapper*, *TreeMix*) retrieve allele frequency data of several populations to efficiently build a comprehensive tree comprising the complex evolutionary history of populations (Pickrell and Pritchard, 2012; Lipson et al., 2013; Harris and DeGiorgio, 2017; Wangkumhang and Hellenthal, 2018; Pathak, 2020). *qpGraph* (Patterson et al., 2012) utilizes the genotype data and proposed tree topology of given mixed and non-mixed populations, adjusting the

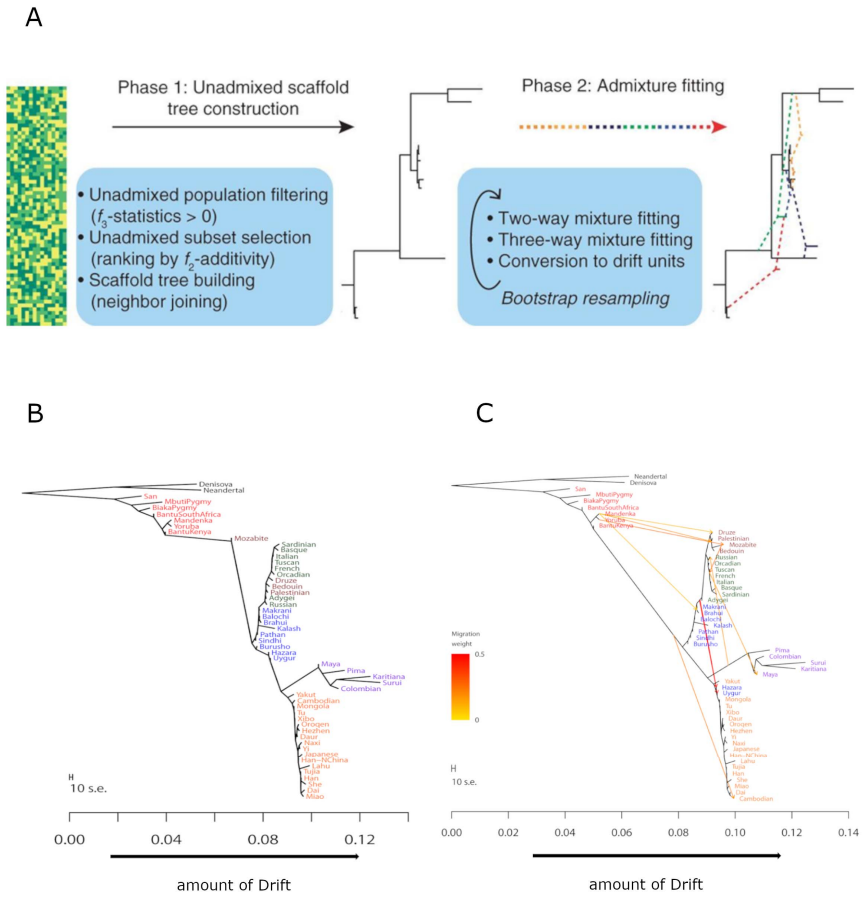

Figure 5. Inferred human phylogenetic tree fitting admixture events.

Graph fitting mixture parameters by *MixMapper*: **(A)** this two-phase method initially builds a tree of unadmixed populations and then tries to fit the remaining populations as admixtures. In the first phase, MixMapper yields a ranking of likely unadmixed trees in order of deviation from f_2-additivity, then this list is used to select a tree as a scaffold. In the final phase, *MixMapper* attempts fitting remaining populations as two- or three-way mixtures between branches of the unadmixed tree. In each case, *MixMapper* applies bootstrap resampling to obtain collective predictions, thus enabling confidence estimation for inferred results. *TreeMix* plots comprise archaic hominins and modern human population groups: **(B)** maximum likelihood tree relating all modern humans and archaic hominins without considering any migration, **(C)** tree allowing for ten migration events between different continental groups of modern humans. The color of migration edges reflects the migration weight, while horizontal branch lengths correlate to the corresponding degree of genetic drifts each branch has faced. Adapted with permission from figure 3, 4 of Pickrell JK, Pritchard JK (2012) published in *PLoS Genet* under terms of Creative Commons Attribution License, and from the publisher "Oxford University Press" for Figure 1 of Lipson et al. (2013) published in *Mol Bio Evol*.

mixture events onto a given tree, computing branch lengths and mixture proportions that best fit the distance metric (*f*-statistics). Though *qpGraph* can integrate manifold admixture episodes, the precondition to specify the graph connecting the populations beforehand is its limitation (Mathieson et al., 2018). Resultantly, *qpGraph* is suggested to be applied with a small number of populations. While *MixMapper* (Lipson et al., 2013), an extension of *qpGraph*, can analyze a large number of populations. It works in two stages **(Figure 5A);** first, it applies f_3-statistic to identify mixed and non-mixed populations building a tree outline of non-mixed populations. Later, it adds one or two mixed populations onto this scaffold. *MixMapper* does not have the obligation of specifying an anticipated tree topology and pre-defined population relationships, both essential prerequisites for *qpGraph* (Lipson et al., 2013; Harris and DeGiorgio, 2017). However, the incapability to support large models with more episodes of admixture is its limitation (Lipson, 2020). The most automated tree-based tool is *TreeMix* (Pickrell and Pritchard, 2012) **(Figure 5B/C)** that needs a user-defined set of populations and number of admixture events to provide a single best-fitting model ((Pickrell and Pritchard, 2012; Lipson, 2020). *TreeMix* is based on a maximum-likelihood framework wherein; first, a branching tree relating a large set of populations is inferred, and later admixture events are added reflecting links between branches (Pickrell and Pritchard, 2012; Harris and DeGiorgio, 2017). If a branch has the support of weight (> 0.5), it is drawn as a tree branch, and in the case of the weight < 0.5, it is migration edge. *TreeMix* relates observed allele frequencies amongst populations using a multivariate normal distribution. The automated approach of building models makes *TreeMix* advantageous, primarily when the prior history of populations is slightly known. However, choosing the appropriate number of admixture events can be difficult in case of unknown complex genetic relationships among a large set of populations (Wangkumhang and Hellenthal, 2018; Lipson, 2020).

Therefore, choosing the best method to create a graph of populations of interest relies mainly on the past knowledge and assumptions of those populations' demographic complexities because each method might be more suitable to a specific context (Lipson et al., 2013).

2.4. South Asia in general

South Asia, the part of the continent enclosed by West, Central, East, and Southeast Asia, is the world's most populous region. The Indian subcontinent (Greater India or Southern Asia) is protected by different geographic barriers such as the Himalayas in the north and east, the Indian Ocean in the south, and Iranian Plateau in the west. The geographical region of South Asia comprises present-day political entities of Sri Lanka, Maldives, Bangladesh, Bhutan, Nepal, Pakistan, and India (UNSD 1999; Petraglia and Allchin, 2007). South Asia can be characterized by three regional natural features – the Himalaya range, Indo-Gangetic Plains, and Deccan plateau. The Himalaya range dominates the majority of North

India, including Nepal and Bhutan. The Indo-Gangetic Plains comprising alluvial plains of the Indus, Ganga, and Brahmaputra rivers, run parallel to the Himalayas encompassing most of North and East India apart from Pakistan and Bangladesh. While the Deccan plateau, located between the Western and Eastern Ghats mountain ranges, runs through West, Central, and South India.

India plausibly served as a major pathway for human migrations consequently, receiving one of the early migrant groups of anatomically modern humans (AMH), when the AMH ventured out-of-Africa to populate the entire world **(Figure 6)** (Kennedy et al., 1987; Cavalli-Sforza et al., 1994; Lahr and Foley, 1994; Quintana-Murci et al., 1999; Misra, 2001a; Thangaraj et al., 2005; Petraglia and Allchin, 2007). This hypothesis has been strengthened by studies suggesting the presence of deep autochthonous ancestries and archaeological artifacts in South Asia (Kennedy et al., 1987; Kivisild et al., 1999, 2003; Basu et al., 2003; Metspalu et al., 2004; Palanichamy et al., 2004; Mellars, 2006b; Thangaraj et al., 2006; Petraglia and Allchin, 2007; Oppenheimer, 2012a, 2012b; Mellars et al., 2013). Later, multiple episodes of people's movements turned India into a blending pot of different genetic components (Majumder and Basu, 2015). Different castes, tribes, and religious societies of South Asia, which have been coexisting for hundreds or thousands of years, differ in their cultural and subsistence practices and even phenotypically and linguistically. In contrast, a fundamental similarity in ritual customs of people from different regions can be a typical example of a rough harmony of diverse religious views, evident in South Asia since long ago (Chaubey et al., 2007). Henceforth, the South Asian region is worth serving as a classic example to investigate the human diversity designed by genetic discrepancies.

Two alternative routes have been proposed for human migrations out of Africa, the "southern route" and the "northern route". The "southern route" proposes that humans moved out of the African Horn crossing the Red Sea, and arrived in Southeast Asia and Australia, traveling along the Indian coastline (Oppenheimer, 2012a). While the "northern route" advocates that humans left Africa through the Nile Valley and kept moving into Central Asia and beyond, including India and other geographical regions. Genetic studies have a perplexing view about the southern and northern route dispersal hypotheses – although the studies on haploid genetic data have been majorly supportive to the "southern route" (Mirazón Lahr and Foley, 1998; Macaulay et al., 2005; Thangaraj et al., 2005; Richards et al., 2006). The archaeological evidence has been meager, probably due to the submerged coastal regions that could have been used for early dispersal by modern humans. On the other hand, a whole genome-based study (Pagani et al., 2015) showed the Northeast African indigenous haplotypes of Egyptians are closer to non-African haplotypes, thus indicating a lesser genetic contribution to contemporary Eurasia via the southern route (Pagani et al., 2015; Molinaro and Pagani, 2018). Nevertheless, the introgression of Neanderthals into genomes of all people living out of Africa endorses the "northern route" dispersal, given the current availability of Neanderthal remains mainly in the Levant region rather than the southern Arabian Peninsula (Krause et al., 2007; Fu et al., 2014).

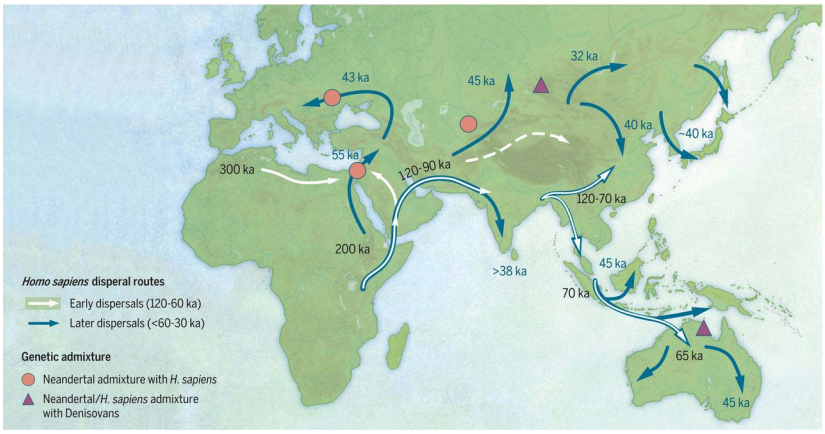

Figure 6. Early and later modern human dispersal routes during Late Pleistocene.

Map of sites with ages, and early and later modern human dispersal routes across Asia. Regions of assumed genetic admixture are also shown. Reprinted with permission from "AAAS" from Figure 1 in Bae et al. 2017 published in *Science.*

Recent advances in computational, statistical, and molecular tools have allowed scholars to establish the modern humans that expanded Out of Africa (OoA) around 50 KYA, as the founder of almost all present-day human populations living outside Africa (Malaspinas et al., 2016; Mallick et al., 2016; Mondal et al., 2016; Pagani et al., 2016). Nevertheless, a minor (~2%) but distinct genetic components in Papuans suggest a possible ancestral contribution from the early dispersal (100–120KYA) of modern humans **(Figure 6)** out of Africa (Pagani et al., 2016). Interestingly, the archaeological traces of Middle Palaeolithic hominins from South Asia gives scholars hope to find a trace of plausible archaic admixture by exploring the massive and unique genetic landscape of the Indian subcontinent (Petraglia and Allchin, 2007; Akhilesh et al., 2018; Clarkson et al., 2020).

2.4.1. Salient features of South Asia

South Asia is one of the most densely populated geographical regions, inhabited by one-fourth of the world population, with more than 664 languages and dialects (Eberhard et al., 2019). The tremendous amount of cultures and ecologies (or natural resources) owned by contemporary India is characterized by its diverse inhabitants, comprised of around 700 tribes (INDIA, 2011) and more than 4000 caste and sub-caste groups that belong to the four primary language families – Indo-European (IE), Dravidian, Austroasiatic (AA) and Tibeto-Burman (TB) (Petraglia and Allchin, 2007). Here, I briefly describe a few of the typical features that plausibly have designed the genetic attributes and extant diversity of peoples in the Indian subcontinent.

2.4.1.1. Climate and Landscapes

The climate of densely populated South Asia generally varies from tropical in the south to temperate in the north. Diverse landscapes range from low-lying regions to high mountains, marshy lands to the estuary, fertile plains to dry plateaus, enduring jungle to sand hills, arid lands, and coastal seashores (Field et al., 2007) influence the climate greatly, giving refuse to diverse flora and fauna. Additionally, plenty of widespread rivers all through South Asia play a crucial role in feeding the people in fertile plains and serve as a vital factor for migration and bi-directional dispersal of people between possible borders (Field et al., 2007). Such diverse ecological conditions might have been crucial in designing the incomparable patterns of migration and adaptation among South Asian populations. Which, consequently, lead to an unparalleled genetic diversity (outside of Africa) in South Asia (Basu et al., 2003; Kivisild et al., 2003; Petraglia and Allchin, 2007; Petraglia et al., 2012; Sengupta et al., 2016; Bae et al., 2017; Wedage et al., 2019a, 2019b).

2.4.1.2. Languages

Amongst the four major language families – the IE, Dravidian, AA, and TB – spoken by South Asian populations **(Figure 7),** the majority of people (~78% and ~20%, respectively) speak IE and Dravidian **(**Petraglia and Allchin, 2007). Whereas the remaining ~2% of populations either speak AA, TB, or minor language families like Tai-Kadai, etc. or fall in the category of linguistic isolates (Moseley, 2008). Previously, the origin of the Dravidian language was linked to proto-Elamo-Dravidian language, spoken by the presently non-existent Elamite group of Southwest Iran, who presumably expanded to the Indus Valley of South Asia with farmers, spreading the early form of Dravidian language (Cavalli-Sforza et al., 1994; Cavalli-Sforza, 1996; Renfrew, 1996; Blazek, 1999).

Later, a pioneer study analyzed the agriculture dispersion and concomitant vocabulary to propose the local origin for the Dravidian language (Fuller, 2003). Whereas people with AA, TB, and IE linguistic affiliations probably derived recently due to the introgression of populations who immigrated to the subcontinent lately (Diamond and Bellwood, 2003). Indo-European speakers typically dwell in North, Northwest, West, and Central India. Dravidian language speakers are mostly limited to four states of South India, but a few populations are exceptionally expanded to other regions of South Asia. For example, Brahui resides in Pakistan, while Gondi, Malto, and Kurukh branches of the Dravidian family inhabit Central and East India and Nepal. Austroasiatic people are scattered mainly in small pockets around people with different language affiliations in India's central and east parts. In comparison, Tibeto-Burman speakers are distributed all along the Himalayan foothills and deep into the northeast part of India. Interestingly, a few tribes living in the Western Ghats and Andaman-Nicobar Islands, referred to as linguistic isolates, do not speak any of the four major language categories; instead, they speak isolated Nihali and Andamanese languages, respectively (Abbi, 2006).

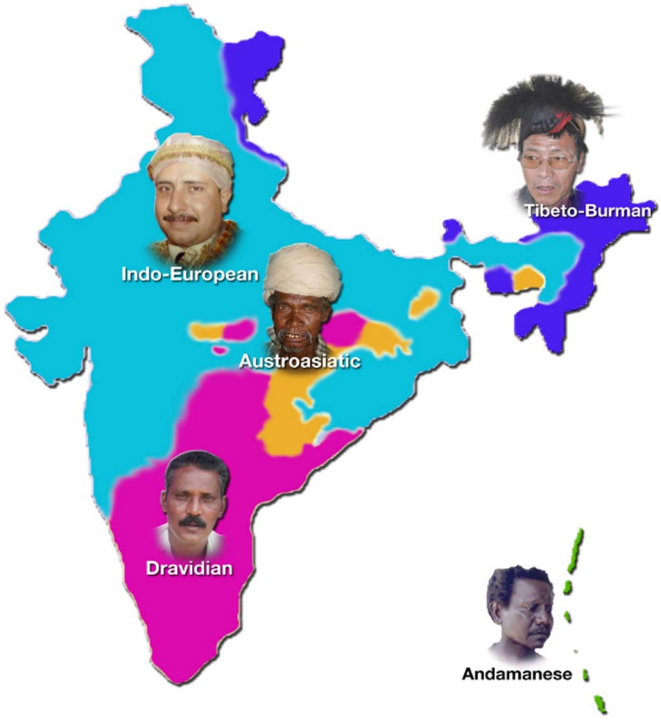

Figure 7. Language map of India.

Geographical distribution of language groups in India. Reprinted with permission of "BioMed Central Ltd." from Figure 1 of Tamang and Thangaraj (2012) in *Investig Genet* under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0).

2.4.1.3. The Social and Caste system

South Asia's extensive and well-structured society is shaped by different layers of caste, tribal and religious groups, cherishing an outstanding level of genetic diversity. The complex Indian social strata are comprised of over 4600 well-described populations and ~700 tribal groups (INDIA, 2011), including several ancestral tribes with an existing hunter-gatherer lifestyle (Tamang and Thangaraj, 2012). The strict endogamy practice is another factor prevalent across all social levels that influences the social structure in South Asia. Endogamy confines the gene flow within caste barriers to preserve their exclusive social/genetic uniqueness, allowing the gene exchange only between population groups of the same social status who share the traditional cultural and ritual customs. A large number of populations follow strict endogamy, forcing several kinds of marriage restrictions across entire ranks of the society and thereby, give origin to diverse population-specific societies who also develop different language dialects.

Although the caste system, an essential part of the Indian socio-cultural system, is typically linked with the Hindu faith, most South Asian populace are either

affiliated to caste groups or tribal groups. The caste system was first documented in the *Rig Veda* (Rao et al., 2009) that was plausibly written during the Iron Age (1700–1100 BC) (Rao et al., 2009; Griffith, 2017). However, the precise dating when *Rig Veda* was written has remained contentious (Misra, 2004; Slavitt, 2015). The *Chaturvarna* doctrine described during the Rig Vedic society categorizes people into four different hierarchical castes underlying their occupational practices; Brahmins (priests and teachers), Kshatriya (warriors and rulers), Vaishyas (Skilled workers and traders), and Shudra (laborers). Additionally, each caste is divided into smaller units of sub-castes that are further split into several clans or Gotras (uninterrupted lineage of the common male ancestor from the ancient time), which are purely exogamous and evade consanguineous marriages. However, consanguineous marriages happen in a few South Indian caste populations (Krishnamoorthy and Audinarayana, 2001; Bittles, 2002; Reich et al., 2009; Nakatsuka et al., 2017). Such dissection of castes into smaller units of sub-castes and clans is the causal factor for the existent diversity and complexity of Indian society (Watkins et al., 2008). However, the ruling authorities of the recent past have been changing the definition of caste/tribe groups consistently.

Such a structured and classified Indian society must have a considerable impact on the extant genetic diversity of modern Indians by separating them into numerous pouches of endogamous groups. Several genetic studies have explored uniparental markers to investigate the caste system of the subcontinent (Kivisild et al., 1999; Bamshad et al., 2001; Quintana-Murci et al., 2001; Wells et al., 2001; Basu et al., 2003; Kivisild et al., 2003; Cordaux et al., 2004a, 2004b; Metspalu et al., 2004; Sahoo et al., 2006; Sengupta et al., 2006; Thangaraj et al., 2006; Chaubey et al., 2007, 2008b). Many of them instigated that a male-mediated voyage of Indo-Aryans primarily created the Indian caste system, which invaded the Indian subcontinent and pushed the native Dravidian speakers to Southern India and Sri Lanka. They also suggested that Indo-Aryans later reputed themselves as the upper caste. It was also added that caste populations vary considerably from tribal groups, and the caste groups are closer to Central Asians and Europeans (Bamshad et al., 1998, 2001; Cordaux et al., 2004a). Though these studies had shortcomings of limited sampling and low-resolution data, they still managed to set the trend to investigate the truth of the Indian caste system. However, a study later reviewed various limitations of the research related to the Indian caste system (Boivin, 2007). Moreover, a study (Mondal et al., 2017) exploited the chrY sequences from different tribal (including Andamanese Jarwa and Onge) and non-tribal groups pointing the complex ancestry of Indian populations that is hard to explain with a single expansion model. The study observed the close genetic affinity between populations of India and southern Europe ans suggested it to be the result of a shared ancient genetic heritage earlier than the Indo-European migration into India.

Later, a couple of studies based on high-resolution genome-wide data revealed that gene flow from West Eurasia had happened before the invasion of Indo-Aryans (Reich et al., 2009; Metspalu et al., 2011). Nevertheless, there is a consensus over the sharing of common paternal and maternal ancestry from the

Late Pleistocene for different castes and tribal groups (Kivisild et al., 2003; Metspalu et al., 2004; Sahoo et al., 2006; Thangaraj et al., 2006; Chaubey et al., 2007, 2008b, 2008a). Successively, additional gene flow occurred from East and West Eurasia (Metspalu et al., 2004; Chaubey et al., 2007).

Observations of genetic studies presented a striking angle to the development and maintenance of the Indian caste system (Thanseem et al., 2006; Chaubey et al., 2007, 2008b). At present, four comprehensive caste groups divide Indian populations into "higher", "middle", and "lower" castes, based on the order of their social ranking (Tamang and Thangaraj, 2012). Sub-tribes born during the upliftment process of ancestral tribal communities might have moved and established themselves among lower caste groups, which in turn may have moved to the next rank and established themselves as a group of middle castes with further empowerments. With more technical and economic advancement over time, these middle caste groups might have subsequently achieved the status of the upper castes. Thus a complete caste system evolved, making the profession of a person symbolic of his caste. Moreover, the caste system also engrossed different episodic migrations from other regions to India into distinct hierarchical levels of the society, sometimes referred to as religious groups (Tamang and Thangaraj, 2012).

2.4.2. Northwest India: the root of ancient Harappan Culture and Vedic Kurukshetra

2.4.2.1. Ancient Harappa

The Indus Valley Civilization (IVC), often also called as Harappan Civilization due to the finding of the first archaeological site at Harappa (Pakistan), is one of the sophisticated societies of the Indian subcontinent that pushed back the outlying ancient times of South Asia by ~2000 years (Shinde et al., 2006; Gangal et al., 2010). The IVC was one of the three supreme prehistoric civilizations of the world that thrived during the early to mid-third millennium Before the Common Era (BCE) (Kenoyer, 1998; Possehl, 2002; Wright, 2010). Emergent archaeological evidence put the IVC forward as the largest Bronze Age civilization, plausibly encompassing an area of ~1 million square kilometers (Possehl, 2002; Gangal et al., 2010). It is also proposed that at the height of the IVC, it had around 5 million inhabitants (McIntosh, 2008).

The IVC **(Figure 8)** that geographically mainly ranged from current Pakistan to the northwest region of India flourished in the Indus River basin (Lawler, 2008; Joseph, 2017). The basin runs across Pakistan and along with a group of perennial rivers that was perhaps monsoon-fed and once flowed near the Ghaggar-Hakra river – sometimes referred to as the mythical Saraswati river of the Rig Vedic era – in Northwest India and East Pakistan (Ghose et al., 1979; Possehl, 1999; Wright, 2010; Giosan et al., 2012a, 2013; Valdiya, 2017). Owing to the similar topographical location to the Ghaggar river, the legendary Saraswati river has primarily been posited as now dried up constant streams of the seasonal Ghaggar river that

coursed through the middle of IVC in the northwest part of India (Ghose et al., 1979; Possehl, 1999; Valdiya, 2017). The majority of IVC settlements were traced to the vast plain, positioned between the Indus and Ganges rivers (the area around the proposed waterways of the Saraswati), and only a minuscule of settlements was located by the Indus river (Singh, 1971; Mughal, 1997; Kenoyer, 1998; Gaur and Vora, 1999; Radhakrishna, 1999; Misra, 2001a; Possehl, 2002; Gangal et al., 2010). Later, a study suggests the geographical borders of IVC covered a massive zone that ranged across southeastern Afghanistan and Pakistan on the west side of the Indus river and northwestern and western states of India on the east of Indus (Shinde, 2016). Since water accessibility is a substantial factor for any urban society to flourish, the presence of a constantly flowing Ghaggar-Hakra (or Saraswati) river at the time of thriving IVC is a general assumption (Radhakrishna, 1999; Wright, 2010; Valdiya, 2017). This assumption is corroborated by the satellite images providing evidence of the Saraswati (or Ghaggar-Hakra) paleo-channel and the unified drainage system in Northwestern India (Pal et al., 1980; Kar and Ghose, 1984; Rajani and Rajawat, 2011; Mehdi et al., 2016). Traces of many mature and late Harappan period settlements also endorse the assumption (Wright, 2010; Petrie et al., 2017). Timing and cause of drying up of Saraswati river have been debatable. A few studies posited the disengagement of the Saraswati river from its enduring glacier sources in the Himalaya (Clift et al., 2012; Giosan et al., 2012b; Singh et al., 2017; Khan and Sinha, 2019) conceivably led to the drying of Northwestern Indian paleo-channels, likely before the establishment of IVC. Whereas other studies proposed the steadying terrain, caused by the absence of uninterrupted flowing river, supported the flourishing of the IVC (Petrie et al., 2017; Singh et al., 2017). Lately, a study tried to resolve the perplexing issues connected to the reality of the Saraswati river – they examined the Holocene period nature of the Ghaggar-Hakra river system by reviewing the temporal change in sediment sources along a particular stretch of the river basin (Chatterjee et al., 2019). The study established the presence of an uninterrupted flowing river that was getting sediments from the Himalayas during 80–20 KYA and 9–4.5 KYA (early phase of IVC), suggesting a significant role of the Saraswati river in the early phase of Harappan establishments (Chatterjee et al., 2019).

The IVC that prospered largely during ~2600–1900 BCE is sub-categorized into three phases indicating start, expansion, and fall of the culture: Early Harappan (3300–2600 BCE), Mature Harappan (2600–1900 BCE), and Late Harappan (1900–1700 BCE) phase (Wright, 2010). The Early Harappan phase had symbolic societies composed of early farmers and pastoralists. In contrast, many planned cities with advanced craft and material culture and established trans-Asiatic trade routes represent the flourishing industries in the urbanized Mature Harappan phase. Nevertheless, by the Late Harappan phase, a foremost deurbanization process brought along population dwindling, non-existent elementary facilities, abandoned settlements, and loss of Harappan scripts (Meadow, 1991; Kenoyer and Meadow, 2010; Wright, 2010). However, recent archaeological sites of Farmana, Bhirrana, Girawad, and Rakhigarhi in the Ghaggar basin

suggest that IVC began at ~5500 BCE (8 KYA) (Sarkar et al., 2016; Shinde, 2016). Nevertheless, urbanization of the Indus Valley probably took place with the development of five major regional centers and numerous minor settlements with emblematic structures and material culture (Kenoyer, 1998; Wright, 2010; Petrie, 2013). These main urban centers are Harappa, Mohenjo-Daro, Dholavira, Ganeriwala, and Rakhigarhi **(Figure 8)** (Wright, 2010). Harappa and Mohenjo-Daro are two early discovered Indus Valley settlements, while the Rakhigarhi is the largest urban center of that period (Shinde et al., 2018). Archaeologists have also revealed numerous Harappan cemetery sites over the last century, including Kalibangan (Sharma, 1999), Farmana (Shinde et al., 2011), Harappa (Hemphill, 1991; Kenoyer and Meadow, 2016), Rakhigarhi (Nath et al., 2015), and Sanauli (Sharma et al., 2007). A recent study attempted to recreate the facial looks of ~ 4.5 KY old individuals from Rakhigarhi cemetery, applying computed topography for craniofacial rebuilding (Lee et al., 2019).

Figure 8. Map of ancient Indus Valley settlements.

Geographical distribution of different ancient Indus valley centers across South Asia. Adapted from "Wikipedia" under the terms of the Creative Commons Attribution License (CC BY-SA 3.0).

The neolithic site of Mehargarh (7000–2500 BCE), one of the earliest settlements of IVC, is in Pakistan and reflects the sign of early farming and pastoralism in the Indian subcontinent (Allchin and Allchin, 1997; Jarrige, 1986, 2006; Dixit et al., 2014). Shreds of evidence for indigenous domestication of barley and frequent depiction of Zebu cattle on Indus seals contradict the general consent that cultivation and maintaining livestock in Northwest India resulted from the Neolithic expansion from Near East (Gangal et al., 2014). A recent genetic study that explored the first ancient specimens from IVC sites of India corroborates the indigenous development of agriculture in South Asia (Shinde et al., 2019).

Indus people used numerous hallmarks of advanced technologies like sophisticated sewage and water systems, houses made of baked bricks, and clusters of industrial buildings highlighting the uniqueness of IVC in eminent town planning. Their technical innovation is also signposted in using copper-bronze metallurgical techniques, typical measurement system, finely shaped stone beads, glossy ceramics, and written language (Shinde et al., 2018). Even the world's oldest glass dated ~1700 BCE is reported at Harappa (Kenoyer, 2008). IVC was plausibly a hierarchical society, which is observed in the use of gold, silver, copper, and tin-based ornaments by social elites; conversely, underprivileged people used low-cost jewelry made from terracotta (Lawler, 2008). Nevertheless, the IVC perhaps had a prosperous and harmonious social order, reflected in finding only a few warfare armaments and many childrens' toys (Gupta et al., 2006). Archaeological findings suggest that Indus merchants mastered the monsoon winds and established a thriving trade network with distant Central Asia, Middle East, and Mesopotamia (Possehl, 2002; Gupta et al., 2006; Lawler, 2008). Sumerians probably defined the IVC as Meluhha and used to call Indus Valley people Meluhhaites, as recorded in Mesopotamian texts (Beck and Beck, 2005; Lawler, 2013). The language spoken by the Indus people is unknown, with scholars having different opinions – different studies linked the Harappan language to Indo-Aryan (Parpola, 1988; Renfrew, 1991) and proto-Elamo-Dravidian (McAlpin, 1981; Renfrew, 1996). A few scholars (Rao et al., 2009) compared the patterns of symbols used in Harappan scripts, suggesting they were closer to Sumerian and Old Tamil words. The Harappan script, generally found on curved seals made of imprinted soapstones to stamp pottery and terracotta cakes, is unanimously said to be written from right to left (Shinde, 2016). However, despite several efforts, the Indus script could not be deciphered mainly because of too short writing, comprising roughly five distinct symbols that lack bilingual text (Parpola, 1979, 1986).

The unexpected end of the Harappan Civilization is another enigma for archaeologists. Generally, the fall of IVC is referred to as a sudden collapse; however, pieces of evidence suggest that even though the populations dispersed from the Indus and Ghaggar-Hakra river valleys to the foothills of Himalaya and plains between Ganga-Yamuna, smaller settlements sustained (Misra, 1993; Kenoyer, 1998; Robbins Schug et al., 2013). Nevertheless, what was the causal factor behind the demise of this flourishing civilization, is debatable. Though different theories have been proposed, including the incursion of Indo-Aryans, change in sea-levels, the decline in rains and trade, and drying up of Ghaggar-Hakra a.k.a.

Sarasvati river forcing Harappans to move eastwards to the Gangetic Plains (Saraswat, 1993; Misra, 2001a; Possehl, 2002; Gupta et al., 2006; Lawler, 2008; Gangal et al., 2010; MacDonald, 2011; Giosan et al., 2012a). Besides, another study (Robbins Schug et al., 2013) proposed the spread of infectious diseases as a causal factor to the demise of the Harappan Civilization. Furthermore, a recent study postulates that the gradual drying up of the Indus Valley summer monsoon rain, due to the change in temperature and weather patterns, made farming unbearable near Harappan cities and, thus, forced the Harappans to move to Himalayan foothills with winter monsoon rain and humid conditions (Giosan et al., 2018).

2.4.2.2. Vedic India and ancient Kurukshetra

The period between the end of urbanization of Harappan Culture (~1500–500 BCE) and the beginning of the second urbanization in the Indo-Gangetic Plain region (~600 BCE) is highlighted as the Vedic period in history of the Indian subcontinent (McClish and Olivelle, 2012). However, it was concentrated mainly in the northern region. The name Vedic comes from the Vedas, the texts which are the primary historic source elucidating the life of that period (McClish and Olivelle, 2012). With the matching archaeological records, Vedas help tracing how the Vedic Culture evolved (Stein and Arnold, 2010).

Eastern Indus basin part of early Vedic period extends along the northwest of India, including historical Kurukshetra (Kenoyer, 2006; Benedetti, 2014). Kurukshetra, a city located in the Haryana state of contemporary India, is supposed to be as old as Indian history (Kurukshetra District, 2009) and well-known for being the main center of the Vedic epic Mahabharata. Mahabharata describes Kurukshetra as the land between the Saraswati and Drishadvati (Srinivasan, 2000; Agarwal, 2003; Witzel, 2003). Kurukshetra is described as – "Dharmakshetra" – the land of Dharma or righteousness in the Srimadbhagwad Gita. Even the Rigveda mentions it, making evident that the name Kurukshetra was prevalent even before the great war of Mahabharata (Roy, 2012). The Puranas suggest that the name Kurukshetra for this Northwest Indian region comes after the King Kuru of the Bharata dynasty (Kurukshetra District, 2009). According to historical texts, the "Ten Kings Battle" took place during the Middle Vedic period between Bharata and Puru tribes. The Bharata tribe won, but the tribes made an alliance and merged into the Kuru tribe, who were dominant from around 1200–800 BCE and formed the first political center of the Vedic period with Kurukshetra as the main power center (Witzel, 2016). A few historians also suggest its existence as a center of religion and culture of India from the pre-Aryan era (Muztar, 1978). Pottery, excavated from Kurukshetra and adjoining areas, point to the pre-Harappans and Harappans possibly being the first settlers of the Kurukshetra region (Bhan, 1972; Kumar, 1978). Late Harappan material culture (c. 1700–1500 BC) is also found in Kurukshetra, fortifying their connection with ancient Harappans (Kumar, 1978). Kurukshetra and adjoining Northwest India has been suggested as the homeland of various ethnic groups like the Kamboj, Gujjar, Ror, and Jat, inhabiting the

region since ancient time and noted in various Vedic and Hindu scriptures (Chidbhavananda, 1965; Purana, 1971; Sitaramiah and Sītārāmayya, 1972; Lal, 1980; O'Flaherty, 1981; Dange, 1984).

Early genetic studies of the Indian subcontinent based on uniparental makers suggest that present-day people in India are mainly comprised of three different genetic components – the West Eurasian, the South Asian, and the East Asian (Quintana-Murci et al., 2004; McElreavey and Quintana-Murci, 2005). Studies based on mtDNA and chrY suggest that Northwest Indian ethnic groups (east of Indus) are genetically intermediate between Pakistani (west of Indus) and North Indian groups (Quintana-Murci et al., 2004). Nonetheless, populations residing west of the Indus region have prevalent West Eurasian maternal ancestry compared to populations living east of Indus with predominantly South Asian maternal lineages.

Therefore, Northwest India needs attention from multidisciplinary research to shed light on the complex structuring of the Indus Valley and Kurukshetra region of Vedic India. To this end, exploring the genetic perspective of various ethnic groups with a long record of presence in the region may prove helpful (discussed in Ref I). Though the Indus Valley region has been a point of attraction for archaeologists, new evidence about this early civilization has emerged via multi-disciplinary approaches. For example, exploring a few first aDNA samples from Indus Valley and surrounding regions of South Asia (Narasimhan et al., 2019; Shinde et al., 2019) has already rejected the Steppe-related genetic ancestry in ancient IVC people. One of them (Shinde et al., 2019) even suggested that the ancient Iranian-related ancestry of South Asians comes from a group that lived before the Neolithic Iranian farmers split from Mesolithic Iranian hunter-gatherers. However, the growing amount of high-coverage data of ancient and modern populations of the region will surely give better insight into the ancient Harappan culture and people in the future.

2.5. Archaeological attributes of South Asia

South Asia has long been a region of tremendous environmental, cultural and human diversity. The genesis of this tremendous diversity has been a challenge for various disciplines over a long time. Another major challenge in settling the pre-history of South Asia is the scarce obtainability of fossils due to adverse environmental conditions (James et al., 2005). Thus, much of the knowledge comes from studying stone tools and other palaeoarchaeological evidence of the subcontinent's early settlements. However, since stone tools were made over the majority of the prehistorical period by the modern humans *(H. sapiens)* and *Homo erectus (H. erectus)* like hominins as well, it is, in fact, hard to distinguish who made which tool (Korisettar, 2017).

2.5.1. Evidence from fossil records

How *H. sapiens* or AMH originated and spread across the globe is a popular research subject, shaping a consensus that AMH evolved in Africa. Such a study compared genomes of hominins and modern humans, suggesting a possible origin for *H. sapiens* about 500 KYA in Africa (Meyer et al., 2016; Stringer, 2016). Studies also observed the earliest existence of morphologically diverse fossils representing *H. sapiens* across Africa, such as ~315 KY old one at Jebel Irhoud in Morocco (Hublin et al., 2017), ~260 KY old Florisbad relic in South Africa, ~195 KY old Omo Kibish relic (McDougall et al., 2005) and ~165 KY Herto fossil from Ethiopia (White et al. 2003). Since both the Jebel Irhoud and Florisbad fossils fall out of the morphological variation of modern humans, they are called early modern humans. Surprisingly, an early modern human relic was recently found in Greece dated ~210 KYA (Harvati et al., 2019), the oldest recorded modern human in Europe and Eurasia (Delson, 2019).

Nonetheless, genetic and palaeoarchaeological records indicate the first appearance of AMH (Omo Kibish and Herto) in East Africa at around 195–160 KYA (White et al., 2003; McDougall et al., 2005; Tishkoff et al., 2009). The broad consensus view suggests that around ~60 KYA, some moved out of Africa into Asia (Mellars, 2006b; Stoneking and Krause, 2011; Kaifu et al., 2014). However, their adventurous journey did not end there, and they kept moving further along the coastal South and Southeast Asia rapidly, finally arriving in Australasia (Macaulay et al., 2005; Mellars, 2006a; Oppenheimer, 2009). Nevertheless, new paleoanthropological evidence indicates the presence of early *H. sapiens* at the Misliya site of Levant by 190 KYA (Hershkovitz et al., 2018b) and in Skhul and Qafzeh between ~130 and ~90 KYA (Grün et al., 2005), after their emergence in Africa during late Middle Pleistocene (Stringer, 2016). Recently, *H. sapiens* traces were also recovered in Arabia ~85 KYA (Groucutt et al., 2018) and Australia by ~65 KYA (Clarkson et al., 2017), demonstrating their possible dispersal before 60 KYA. The archaeological and paleoanthropological findings suggest that AMH plausibly migrated out of Africa several times at the beginning of the Late Pleistocene. This is congruous to the recent genetic evidence that early humans conceivably also expanded across Southern Asia before the migration that took off at ~60 KYA (Pagani et al., 2016) and inscribed in our shared genome (Groucutt et al., 2015; Bae et al., 2017). However, the dispersal route, taken by the earliest modern humans to the Indian subcontinent, and time of arrival is a highly debated topic (Mellars, 2006b, 2006a, 2006c; Petraglia et al., 2007, 2009; Clarkson et al., 2009, 2012; Mellars et al., 2013; Mishra et al., 2013), primarily due to scarce relics of human fossils from the Indian subcontinent.

The growing recovery of fossils from sites in East and Southeast Asia, too, supports the early influx of *H. sapiens* across Southern Asia during the early phase of the Late Pleistocene. Several fossil teeth dated between ~130–70 KYA, extracted from Fuyan Cave (Liu et al., 2015) and Luna cave in China (Bae et al., 2014), probably represent early *H. sapiens* (Michel et al., 2016). 70–46 KY old *H. sapiens* fossils recovered from Laos (Demeter et al., 2017) and ~68 KY old

teeth accompanied by ~75 KY old faunal evidence from the site of Lida Ajer cave in Sumatra (Westaway et al., 2017) corroborate the presence of early *H. sapiens* in Southern Asia. On the contrary, ancient humans of India are understudied due to harsh environmental conditions. Only a couple of fossils related to hominins are recovered in the Indian subcontinent. Even the archaeological proof suggestive of the presence of modern humans in South Asia earlier than 50 KYA is scarce (Haslam et al., 2010; deMenocal and Stringer, 2016; Korisettar, 2017). In the Indian subcontinent, the earliest and most precise records of *H. sapiens* come from Sri Lanka, where partial skeletons recovered from Fa-Hien cave and Batadomba-lena Cave dated between ~31–28 KYA (Kennedy and Deraniyagala, 1989; Deraniyagala, 1992). In India, 17–10 KYA old *H. sapiens* skeletons were recovered from Jwalapuram (JW 9) site in Jurreru Valley (Clarkson et al., 2009). Though technological advances enabled the analysis of genetic materials from a few Bronze Age to early historic period fossils in parts of South Asia (Narasimhan et al., 2018; Shinde et al., 2019), the temperate and dry environment of the subcontinent has been a limiting factor to obtain fossil records of the pre-Neolithic period.

The only known hominin fossil in South Asia that has disputed identity and best described as the intermediate form of *Homo* named *Homo sp indet.* (Athreya, 2007), was excavated from the Hathnora site in the Narmada river valley (Sonakia and Kennedy, 1985). Few scholars (Kennedy, 2001; Cameron et al., 2004) suggest Hathnora hominin remains are associated with late Middle Pleistocene (~236 KYA), whereas others suggest (Patnaik et al., 2009) a recent date of 160–85 KYA. Other fossils in the close-by sites are not clearly dated (Sankhyan, 1997a, 1997b, 2005; Sankhyan et al., 2012), but the recovery of Netankheri remains from below the Youngest Toba ash suggests they existed earlier than 75 KYA. Nevertheless, these fossils highlight the significance of the Narmada Valley in the subcontinent's early human settlements. Additionally, most rock paintings and engraving records are located in Central Indian corridor between Narmada and Son Valley, signifying the region's importance in the context of early human colonization (Misra, 2001a; Chakrabarty, 2009; Witzel, 2012; Cummings et al., 2014). These rock paintings and engravings depict humans and animals on walls of rock shelters and caves that were occupied mainly by Upper Paleolithic and Mesolithic people, e.g., the Central Indian Bhimbetka rock-shelters house cave paintings of the Mesolithic period (Misra, 2001a; ASI, 2003).

2.5.2. Evidence from Paleolithic and Microlithic tools

Early hominins (archaic hominins) who ventured out of Africa used bulky stone hand-axes and hatchets, characterizing the Acheulian culture. However, *H. sapiens,* who spread out later, were equipped with advanced stone tools comprising superior cutting tools (blade technology) and spearheads made of small stone flakes during Middle Paleolithic culture. Therefore, the period between the end of Lower Paleolithic Acheulian culture until the advent of Middle Paleolithic

(or Middle Stone Age) ~300–200 KYA (Tryon, 2003, 2006; James et al., 2005; Porat et al., 2010; Sahle et al., 2014; Álvarez-Alonso, 2014; Richter et al., 2017), was a transition phase of noticeable changes in hominin behavior. Conceivably, choosing the advanced stone flake-based tools over the inferior Acheulian culture stone tools of early hominins was the primary reason for the quick expansion of *H. sapiens* across Eurasia (Akhilesh et al., 2018). Researchers have been trying to investigate this intriguing factor by dating the stone tools recovered around the world and testing whether such a transition of tools did happen in different geographical regions or not.

In India, the earliest Acheulian tools come from Attirapakkam in Tamilnadu dated >1 MYA (Pappu et al., 2011). Later, during the 800–200 KYA period, stone tools are abundant in the Siwaliks, Vindhyan basin, and Chhota Nagpur regions of Central India, and the South Indian Bhima, Kaladgi, Cuddapah, and Kortallayar basins as well. These stone tools believably were made by early hominins (Korisettar, 2017). Middle Paleolithic tools were observed only after 140 KYA (Pappu et al., 2011). However, a recent discovery of the sleek and younger stone flakes at Attirapakkam pushed back the presence of Middle Paleolithic tools in India that possibly started at ~385 KYA and continued until ~172 KYA (Akhilesh et al., 2018). Strikingly, this suggests that Middle Paleolithic records in India were contemporaneous to those of Africa and Europe (McBrearty, 2003; Tryon, 2003, 2006; James et al., 2005; Hovers and Kuhn, 2007; Sharon, 2007; Porat et al., 2010; Armitage et al., 2011; Adler et al., 2014; Sahle et al., 2014; Álvarez-Alonso, 2014; Richter et al., 2017). This fact put forward many unanswered questions and skepticism over the earlier consensus view of local development of Middle Paleolithic tools in India (James et al., 2005; Clarkson et al., 2009). Such a severe and abrupt tool transition at Attirapakkam agrees with the debated hypothesis of human migration out of Africa earlier than 60 KYA, too.

The excavations at Jwalapuram in the Jurreru River valley of Andhra Pradesh revealed stone tools interlaced between volcanic ash layers produced by Toba eruption (Youngest Toba Tuff-YTT) of Indonesia ~74 KYA (Petraglia and Allchin, 2007; Petraglia et al., 2007). The Toba eruption (YTT) occurred in the Indonesian Sumatra island (~74KYA) and likely caused a drastic reduction in temperature and global human populations. A lately discovered stone tool industry dated ~79–65 KYA at Dhaba site in the Son River Valley overlaps the incident of Toba eruption (~74KYA). It bears a resemblance to that of Middle Stone Age Africa and Arabia (Clarkson et al., 2020). Also, the study observed a gradual transition from Levallois to microlithic technology ~48 KYA, therefore, substantiating the human occupation of northern India ~74 KYA. Upper Paleolithic tools, characterized by microliths, appeared first at the Mehtakheri site of Vindhya basin dated 45 KYA. Later, more microliths were recovered from Jwalapuram (JWP9) and Patne in Deccan Province, dating earlier than 34 KYA and 27 KYA. Microliths were also found in Sri Lanka from ~38 KYA. Hence, the presence of modern humans in the Indian subcontinent by 40–35 KYA is pretty convincing, especially in the light of the evidence of the persistent occurrence of stone microblades from around 38 KYA at Jwalapuram, which signifies a fast and

systematically progressing microlithic technology throughout South Asia by 35–28 KYA (Clarkson et al., 2009; Petraglia et al., 2009).

Archaeological findings show that *H. sapiens* dispersed out of Africa with Middle Paleolithic tools (Lahr and Foley, 1994; Foley and Lahr, 1997; LAHR and FOLEY, 1998). These tools are technically quite similar to the Middle Paleolithic assemblages of Arabia (85 KYA) and India (78–74 KYA) (Petraglia et al., 2010; Clarkson et al., 2020), suggesting an earlier presence of *H. sapiens* in both regions. *H. sapiens* remains were recently found in Arabia from ~85 KYA (Groucutt et al., 2018) but are missing in India. However, this chronometric age is beyond 70–50 KYA when modern human gene pools coalesce. Many scholars (Petraglia et al., 2010) argue that the absence of *H. sapiens* remains in South Asia perhaps is the consequence of an extensive arid phase across South Asia that might have expanded to the Thar desert. Nevertheless, this aridity also created tropical forests in Sri Lanka and Southwest Asia, which could have been used as a refuge by forager groups. Some of these groups plausibly survived the Toba super-eruption, though there was a population-level impact (Ambrose, 1998; Jones, 2007). Some suggest that the extensive population growth and replacement after ~35 KYA, driven by demographic dynamics during the Middle and Late Holocene period, left the genetic signature of these populations out of the present-day human gene pool (Petraglia and Allchin, 2007).

However, the present uncertainties regarding modern human arrival and routes and many puzzling evolutionary questions suggest that examining the Indian subcontinent via a multidisciplinary method could be significant to reveal the truth behind the arrival events of modern humans in South Asia and further.

2.6. Genetic overview of South Asia

2.6.1. Demographic inferences from haploid data

The South Asian populations exhibit an extraordinary genetic diversity that is just second to that of sub-Saharan Africans (Reich et al., 2009; Xing et al., 2010; Mallick et al., 2016; Pagani et al., 2016). Such a striking genetic diversity is in tandem with clear archaeological (Petraglia et al., 2007, 2010; Haslam et al., 2010) and maternal genetic signals of AMH in South Asia by 50 KYA (Kivisild et al., 1999; Thangaraj et al., 2005, 2006; Mellars et al., 2013). Together, these pieces of evidence advise a likely use of the Indian subcontinent as the earliest passageway for dispersal of AMH through West Asia when they exit out of Africa (Kivisild et al., 1999; Macaulay et al., 2005; Majumder, 2010; Mellars et al., 2013). Maternal (mtDNA) genetic studies mainly echo the earlier demographic processes of the pre-Holocene, including the initial human occupation of the subcontinent at around ~55–65 KYA and key population changes during the later Pleistocene period. However, studies based on paternal Y-chromosome (NRY) tease out the demographic dynamics of South Asia in the Holocene period of the last 10 KYA (Kivisild et al., 2003; Chaubey et al., 2007; Silva et al., 2017). A significant

number of uniparental genetic studies agree that the genetic components of South Asian populations though primarily carry deep-rooted indigenous ancestries, yet also harbor a genetic component shared with West Eurasians (Kivisild et al., 1999, 2003; Basu et al., 2003; Metspalu et al., 2004; Sahoo et al., 2006; Sengupta et al., 2006; Thanseem et al., 2006; Underhill et al., 2015; Mondal et al., 2017; Sahakyan et al., 2017). The frequency of the genetic component shared with West Eurasians demonstrates a gradient from North to South of India, encompassing the highest maternal and paternal West Eurasian ancestries (~50%) in Northwest India that declines toward the South and East of the continent, reaching to around <10% (Metspalu et al., 2018). The low genetic variation observed in mtDNA lines (Bamshad et al., 2001; Roychoudhury et al., 2001; Basu et al., 2003; Kivisild et al., 2003; Metspalu et al., 2004; Chaubey et al., 2007) is contrary to the reasonably high variation in chrY ancestries of Indian populations (Bamshad et al., 2001; Basu et al., 2003; Kivisild et al., 2003; Sengupta et al., 2006; Chaubey et al., 2007). Thus, suggesting a pattern of the higher number of male immigrants than female counterparts (Bamshad et al., 1998) during early to prehistoric India (Majumder and Basu, 2015).

2.6.2. Demographic inferences of South Asia from genome-wide data

The first diploid genome-based study, which included disease candidate genes and single-nucleotide polymorphisms (SNP) assay, was conducted on 405 SNPs from 1871 individuals of 55 diverse Indian populations, revealing their genetic status in-between the Caucasians and Asians (Indian Genome Variation Consortium, 2008). The study also discovered fuzzy geographic clustering amid population groups and found that genetic kinships among Indian populations were determined mainly by their tribal/non-tribal status and language. Another study by the HUGO Pan Asian SNP consortium put forward the genetic sharing of Indian populations to Europeans (Consortium et al., 2009), mostly in line with the expansion of Indo-European speakers.

However, the genome-wide survey of South Asia (Reich et al., 2009), by scanning more than 500K SNPs, observed that the genetic closeness of Indian populations to West Eurasians/Central Asians follows a north to south gradient (or Indian Cline). Wherein the higher caste groups are genetically closer to West Eurasians than the lower caste groups. Furthermore, the study projected most contemporary South Asian populations as a mixture of two purported ancestral populations – Western Eurasian-related Ancestral North Indian (ANI), and Ancestral South Indian (ASI) which is equidistant to both West Eurasians and East Asians. ANI component is higher among North Indians that drops from Northwest India toward South, whereas ASI is higher among South Indians. Another genome-wide study analyzed SNP data from 30 Indian populations and observed a similar pattern; they even tried to date the admixture event between ANI and ASI applying a haplotype diversity-based approach (Metspalu et al., 2011). Moorjani et al. (Moorjani et al., 2013) applied an LD decay-based method

to date the ANI-ASI admixture event aptly and discovered a mixture event nearly 1900 to 4200 years ago. Mainly, extensive admixture occurred among different population groups of India during this period, followed by an apparent shift of populations to endogamy (Moorjani et al., 2013; Basu et al., 2016). A study tried (Nakatsuka et al., 2017) to look into the consequence of endogamy practices by skimming through the genome of 250 diverse ethnic Indian populations. The study anticipated that the endogamy-induced founder population effect is a causal factor for recessive diseases specific to a particular population, making almost one-third of the subcontinental populations prone to such diseases. Additional genome-wide studies from South Asian geography have been added with time (Chaubey et al., 2015, 2016, 2017; Mondal et al., 2016; Tamang et al., 2018; Debortoli et al., 2020). One such study explored the Indian Jews population that migrated to South Asia during historical period, establishing their genetic heritage from West Eurasia and comprehensive intermixing with local Indians (Chaubey et al., 2016). Mondal et al. (Mondal et al., 2016) compared whole-genome sequences of Andamanese individuals with mainland India and other global populations finding a common single origin for all Asian and Pacific populations thus, supporting a single wave of human expansion out of Africa. In another study, Yelmen et al. (Yelmen et al. 2019) applied the local ancestry deconvolution method, revealing differential demographic antiquities and contrary selective burdens in contemporary populations of the subcontinent. Whereas, Kai et al. (Tätte et al., 2019) observed that although the Indian Austroasiatic (AA) populations share higher IBD to the geographically far Malaysian peninsular tribes than Southeast Asian neighbors, their genome is characterized mainly by South Asian and Southeast Asian like components due to the admixture event ~2000–3800 KYA. A recent study (Wall et al., 2019) that attempted to bridge the representation of non-Europeans in genetic studies included 598 sequences of diverse Indian populations gathering new insights into archaic introgressions and the correlation of genetic diversity to caste and language groups.

2.6.3. Demographic inferences of South Asia from aDNA

2.6.3.1. Pre- and early historical demography of South Asia

The ability to examine extinct populations made the aDNA studies a driving field in determining ancient population movements that shaped the current global human genetic variation and even furthering the insights of past human behavior and socioeconomic processes at the local level (Slatkin and Racimo, 2016; Nielsen et al., 2017; Skoglund and Mathieson, 2018; Callaway, 2020; Racimo et al., 2020). A few such studies compared the genomes of ancient West Eurasian specimens and modern South Asians to solve the puzzled demographic history of South Asia (Broushaki et al., 2016; Gallego-Llorente et al., 2016; Lazaridis et al., 2016). They reported that the agriculture in South Asia was spread by ancient Neolithic farmers of Iranian Zagros mountains (Iran_N), unlike a genetically

different Neolithic farmer group from Anatolia (Anatolia_N) that triggered the spread of agriculture in Europe. Lazaridis et al. (Lazaridis et al., 2016) also suggested the ANI-ASI cline displayed by contemporary South Asian populations is a result of admixture of three genetic components that relate to Neolithic Iranian farmers, Yamnaya like Early and Middle Bronze Age Steppe populations (Steppe_EMBA), and South Asia specific Onge group. Despite an explosion in aDNA-based studies from West Eurasia, ancient specimens are scarce in the subcontinent, probably due to the harsh tropical climatic conditions. However, recent scientific advances have ensured to retrieve the DNA material also from ancient specimens in South Asian climatic conditions. A few of the early aDNA studies from South Asia were limited to uniparental genomic data. One such study analyzed the mtDNA markers on the bone relics of Queen Ketavan of Georgia from a Goan church and established the authentic identity of the bones (Rai et al., 2014).

However, first genome-wide aDNA data from the Indian subcontinent, collected from the prehistorical and historical period Swat Valley region of Northern Pakistan (Narasimhan et al., 2018), articulated a considerably elaborate evolutionary past of the Indian subcontinent and Central Asia **(Figure 9)**. The study analyzed 362 new aDNA data that further expanded to 523 (Narasimhan et al., 2019) from the broader South and Central Asian region, indicating the connection of Middle and Late Bronze Age Steppe (Steppe_MLBA) pastoralists, rather than the earlier proposed Yamnaya (Steppe_EMBA), to the South Asian genetic pool and spread of IE language during 2300–1500 BCE. The study redefined the earlier proposed genetic composition of the present-day South Asian populations, linking their genetic heritage to a complex set of geographically and temporally close sources. The Indus Periphery (IP), proposed to be related to the ancestry of ancient IVC people, contains varying proportions of ancestry from Iranian farmers (Iran_N), Ancient Ancestral South Indians (AASI), and Western Siberian Hunter-Gatherers (WS_HG) **(Figure 9b),** and is genetically a prime ancient source population for modern South Asians. Post IVC collapse, IP moved eastward and admixed with AASI, forming ASI component; whereas, in northwestern region, they encountered further incoming Steppe_MLBA, thus by making ANI component **(Figure 9c)**. Nevertheless, a recent study (Shinde et al., 2019) successfully extracted a single Bronze Age period aDNA from the Indian side of IVC settlement, Rakhigarhi, suggesting the Bronze Age Indus individual had predominant ancestry related to Iranian farmer and AASI (Onge) but lacked the Steppe ancestry. Interestingly, the Iranian-related ancestry displayed by the IVC individual was a former group that existed earlier than 12KYA, before the split of Iranian farmers and hunter-gatherers when farming did not begin in Fertile Crescent (FC). Thus, they suggested that agriculture in South Asia is mostly indigenous phenomena that developed through the exchange of ideas instead of migrations of people from FC. The study observed IP individuals genetically identical to the BA Indus Valley individual, clubbing them together for all analyses, and named IVC Cline claimed as the single most source for all present-day Indians (Shinde et al., 2019). Contemporary Indians are suggested to have genetic components from IVC Cline, AASI ancestry, and Steppe_MLBA (Narasimhan et al., 2019;

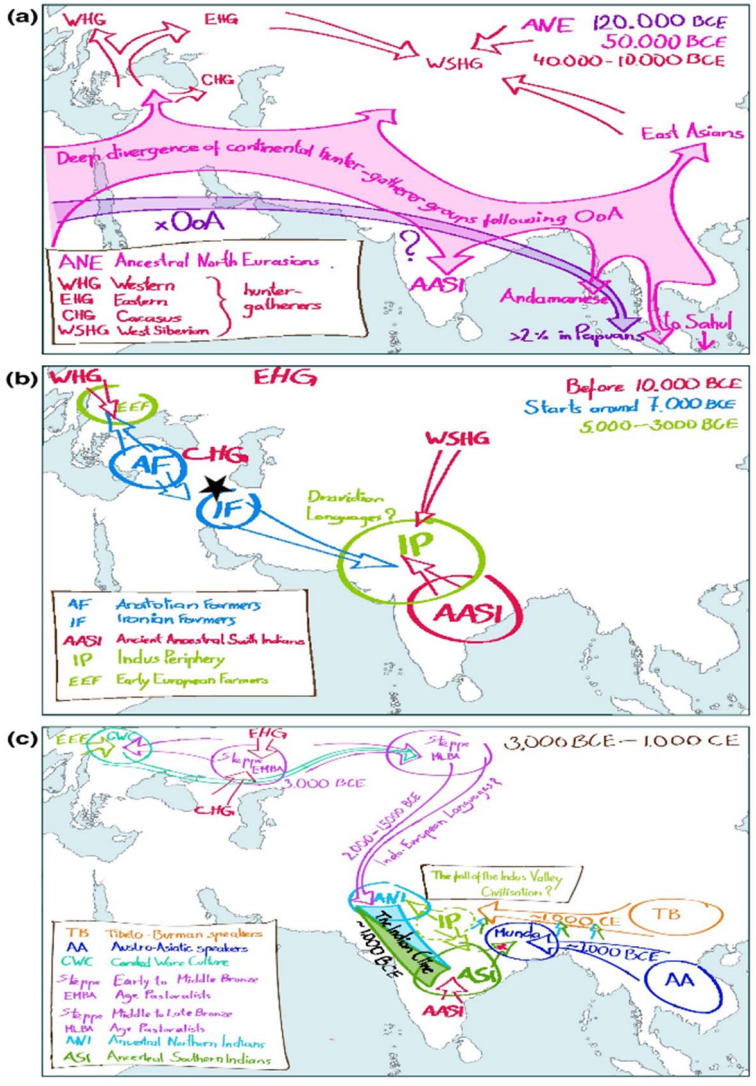

Figure 9. A sketch of the peopling history of South Asia.

Depicting the full complexity of available reconstructions is not attempted. Placing of population labels does not indicate precise geographic location or range of the population in question. Rather the aim is to highlight the essentials of the recent advancements in the field. The scenario is divided into three time horizons: Panels **(a)** before 10 000 BCE (pre agriculture era.); **(b)** 10 000 BCE to 3000 BCE (agriculture era) and **(c)** 3000 BCE to prehistoric era/modern era. (iron age). *The star* in Figure b is an attempt to modify graph to match the latest finding by Shinde et al. 2019 that Iranian related ancestry in ancient and modern South Asians comes from an older (~12 KYA) Iranian hunter-gatherer group, instead of the previously claimed Neolithic farmers of Iran.* Adapted from Figure 1 of Metspalu et al. 2018 published in *Curr Opin in Genet & Develop*, with permission from "Elsevier" publishers.

Shinde et al., 2019). Another study analyzed 38 genome-wide aDNAs and tried to unravel the unknown origin of hundreds of skeletons scattered in the Roopkund Lake of Himalayan mountains (Harney et al., 2019). The majority of skeletons that comprised a genetic component related to contemporary South Asians were dated to ~800 CE. In contrast, a few skeletons with eastern Mediterranean-like genetic ancestry, including the only skeleton with Southeast Asian-related ancestry, were dated to ~1800 CE.

2.6.3.2. Archaic admixture pattern of South Asia

Recent advances in studies on genomic data from worldwide populations conclude that plausibly all non-Africans descend from AMH (Malaspinas et al., 2016; Mallick et al., 2016; Mondal et al., 2016; Pagani et al., 2016) that spread out of Africa to populate rest of the world. During the course, they encountered and admixed several species of hominins that are extinct now (Skoglund and Mathieson, 2018). aDNA has enabled the unprecedented teasing out of such genetic traces left during past interactions between archaic hominins and ancestors of modern humans across different continents, including South Asia. Comparative analysis of Neanderthals' genomes to that of present-day populations revealed Neanderthals' ancestry contribution to non-Africans falling in the range of 2–2.6% (Prüfer et al., 2014, 2017; Browning et al., 2018). In contrast, the genomic analyses with Denisovans' DNA discovered around 5% genetic contribution from Denisovans to the contemporary Oceanians, and probably a little bit to even East and South Asians (Reich et al., 2010; Skoglund and Jakobsson, 2011; Meyer et al., 2012; Qin and Stoneking, 2015; Malaspinas et al., 2016; Browning et al., 2018).

The complex archaic introgression pattern in Asia (Mondal et al., 2019; Teixeira and Cooper, 2019) and the earlier occupation of the region by various hominin groups (Détroit et al., 2019) back the mating of Asian ancestors with unknown archaic hominins. After the first encounter event with Neanderthals, AMH split up into two branches; first populated the geographical region of Europe as early Europeans. The second branch gives rise to a common ancestral group that spread across the vast regions in South, Southeast, and East Asia and Sahul. Recently, a study (Mondal et al., 2019) applied deep learning methods on genetic data to model the demographic dynamics. The study observed a distinct genetic signal of ~3% in Andaman Islanders and Australian Aboriginals, which was contributed by a ghost hominid population separate from the earlier known Neanderthals and Denisovans. The ghost hominid, which is genetically halfway between Neanderthals and Denisovans, probably interbred with the second branch of AMH somewhere in South Asia. Nevertheless, archaeological relics (aforementioned) claim that Middle Paleolithic hominins were in South Asia earlier than the leading expansion of modern humans out of Africa around 50–55 KYA. Altogether, the facts discussed above make diverse genetic backgrounds of South Asians a compelling case that could lead to traces of additional archaic admixture, if any, or many other unknown stories of past interactions between archaic hominins and modern humans.

2.6.4. Languages and genetic correlation in South Asia

Genetically, IE and Dravidian speaking populations (see also ch 2.4.1.2) generally follow a reciprocal cline from north to south and south to north, respectively, formed due to the admixture between two extremes: ANI and ASI (discussed above). ANI is over-represented in IE speakers while ASI predominates in Dravidians. Meanwhile, the geographical origin of AA language speakers, the third-largest group in India, has been a point of numerous discussions among researchers (see, e.g. (Fuller, 2007; Kumar et al., 2007; Chaubey et al., 2011a)). Indian AA language has two main branches. The Munda branch is the most prevalent one; in contrast, the other one, Khasi-Aslian, spoken by a few Meghalaya and Andaman Island groups, has higher local Indian genetic ancestry and is relatively close to Indian TB speakers (Diffloth, 2009; Chaubey et al., 2011a). Lately, it is posited that the Munda language arrived in India around 2500–1500 BCE, when a small group of Southeast Asian Mainland Austroasiatic speakers, who knew rice farming and spoke pre-Proto-Munda language, arrived at the Eastern Coast of India via route around the Bay of Bengal.

These rice farmers then intermixed with the inhabiting Indian populations, resulting in proto-Munda and Munda languages (Rau and Sidwell, 2019). Genetic studies agree that Indian AA speakers (Munda) originated from the westward expanding Southeast Asians, followed by a male-mediated migration from Southeast Asia and interbreeding with local Indian groups later on (Chaubey et al., 2011a; Arunkumar et al., 2015). Using genome data from South and Southeast Asia, a more recent study refined the date of admixture events in different Munda speakers to the range of 2000–3800 years ago (1800–0 BCE) (Tätte et al., 2019). Whereas previously, uniparental data-based admixture was suggested to occur at a somewhat earlier period around 3000–2000 BCE (Arunkumar et al., 2015). Tätte et al. also suggested that ~20% of Munda genetic ancestry comes from Lao like Southeast Asian population; the genome of Munda people is comprised of lesser West Eurasian related ancestry and are more similar to Onge – the South Asian HG (AASI), rather than ASI representing Dravidians, such as Paniya. Interestingly, populations that belong to the South Munda group contain higher Southeast Asian genetic contributions than those of the North Munda group (Tätte et al., 2019).

TB speakers, concentrated mainly in Northeast India and upper North India in Himalayan foothills, indicate their strong genetic heritage from East Asia (Reich et al., 2009; Metspalu et al., 2011; Tamang et al., 2018) rather than South Asia, owing to admixture (in last 1000 years) with westward spreading people from the east of Himalaya (Tamang et al., 2018). Interestingly, sources responsible for the East Asian genetic legacy of TB groups vary among themselves. TB Populations that live in the extreme north of India, near the Tibetan region, share a substantial amount of ancestry to Tibetans than Han. In contrast, Northeast Indians and Southeast Asians commonly display a higher genetic sharing with the East Asian Han population (Tamang et al., 2018).

In general, the genetic architecture of populations in South Asia is linked to their linguistic affiliations. Thus, as a rule, individuals from the same language group are genetically closer – an overwhelming but not at all universal observation globally (Cavalli-Sforza et al., 1994). Consequently, one would expect that the genome of a population that has gone through a recent language transition might reveal this event. However, it is not always true for Indian populations. Indeed, several populations adopt a new language with an almost invisible impact on their genetic composition, thus contradicting the correlation of a population's gene pool to its linguistic affiliation. For example, Gond – the largest Indian tribal group, and Bharia populations speak the Dravidian language; nonetheless, their genetic composition endorses that they share higher ancestry with Indian AA groups than with other linguistic Dravidians (Sharma et al., 2012; Chaubey et al., 2017). Another example is the Musahar population, an IE (Hindi) speaking group living in northern states of India that is genetically closer to Indian AA groups than to their IE-speaking neighbors (Chaubey et al., 2008b). However, the Pakistani Brahui group is a remarkable example of an opposite case; the Brahui population is genetically similar to their geographic neighbors but is the only group in the northwest of the subcontinent that retains the Dravidian language, which is now limited to the southern region of India only (Pagani et al., 2017).

2.6.5. Sex-specific ancestry patterns

Analysis of genome-wide data that holds numerous pieces of genetic information compared to mtDNA/chrY is the preferred approach to construe the human admixture (Wangkumhang and Hellenthal, 2018). Nonetheless, the mtDNA and chrY are tremendously treasured approaches to detect the sex-specific mixing in the case of skewed maternal or paternal genetic contributions from one of the admixing ancestral populations (Jobling et al., 2014). mtDNA and chrY provide an opportunity to identify specific female and male lineage clades (or haplogroups), respectively, and correlate them to a geographical region. This advantage gives the liberty to compare the two systems and detect sex-specific dispersal patterns that can be useful in disclosing a population's remarkable social structure (Karmin et al., 2015). The uniparental genetic data points out significant differences in the heritage of the male and female lines of descent in the Indian Subcontinent (Bamshad et al., 1998; Kivisild et al., 1999, 2003; Basu et al., 2003; Metspalu et al., 2004; Chaubey et al., 2007; Majumder and Basu, 2015; Underhill et al., 2015; Silva et al., 2017). Which is the result of sex-specific gene flows that probably took place within the last 5 KY. The chrY shows a distinctly higher West Eurasian related male lineages ranging from ~ 50–90% across South Asia, contrary to the mtDNA that suggest a much higher presence of autochthonous lineages varying from ~70–90% (Chaubey et al., 2007; Silva et al., 2017). Such a sex-specific pattern is also visible in the East Asian ancestry component of TB and AA speakers; however, it is somewhat less noticeable due to a much lower contribution to respective genetic components (Chaubey et al., 2011a).

2.6.6. Historical immigrants into South Asia and their genetic antiquity

South Asia has gone through a number of prehistorical and historical immigration from various human populations, shaping its modern-day gene pool. Such ethnic groups that arrived in the subcontinent recently, during the historical period, have been a matter of scholarly interest, resulting in genetic studies of a few immigrants like Makranis (Laso-Jadart et al., 2017), Siddis (Shah et al., 2011), Jews (Behar et al., 2010; Chaubey et al., 2016; Waldman et al., 2016a, 2016b), Parsis (al-Maghtheh et al., 1993; Qamar et al., 2002; Mohyuddin and Mehdi, 2005; López et al., 2017), and Muslims (Eaaswarkhanth et al., 2010). Makranis and Siddis communities are two genetically studied migrant groups of recent historical time, which perhaps arrived here owing to the Indian Ocean Slave Trade. Makranis orally suggested to have roots in Ethiopia were transported to Pakistan as part of the Indian Ocean slave trade during the 8th–19th century. A genetic survey of Makrani people revealed their origin ~300 years ago due to the amalgamation of ancestral lines from two ethnic groups, the local Baluch tribes and Bantu-speakers of East and Southeast Africa (Laso-Jadart et al., 2017). The majority of them also found to carry the genetic variants common in African populations that protect against malaria. The Siddis, too, are part of this slave trade, who were sold to Sultans and Nawabs of the Indian subcontinent by Portuguese merchants during the 17th–19th centuries. The genetic pattern of Siddis suggest that although one-third of Siddi genetic ancestry is related to local South Asians, they retain around two-third of their original African related ancestry; this fusion of ancestries is the result of an admixture event during the last 200–300 years (Shah et al., 2011). The extradition of African slaves to Arab and South Asia from East African trading posts set by Arabic traders is widely suggested in several topical studies (Blench, 2014; Hellenthal et al., 2014; Brucato et al., 2017). Indeed, it is posited that the Indian ocean has been used for both ways movement, of people and languages, between populations living in Eastern Indian Ocean rim (including South and Southeast Asia) and Western Indian Ocean rim people (including East Africa and Arabian Peninsula) from early historical time until 19th century (Brucato et al., 2019). Muslim community of the Indian subcontinent, likely descendants of Arabs during the 7th century quest of Arabian empire followed by multiple incursions from Turkish people (Schimmel, 1982), display a genetic pattern differing from the above two African migrants. The genetic studies of Indian Muslims, though observed a minor yet detectable level of genes coming from Iran and Central Asia, did not reveal any significant Arab-related gene flow. Nevertheless, the present-day Muslims across India share most of their genes with their regional neighbors of non-Muslim communities, suggesting that the spread of Islam in South Asia mostly transpired through cultural conversion (Eaaswarkhanth et al., 2009, 2010; Mohsenpour and Chandrasekar, 2019). The other two prominent historical immigrants to South Asia, the Parsi and Jews communities, who migrated from the Middle East region, are discussed in further detail in the coming section keeping their contextual relation to this thesis.

2.7. Parsis and Jews: the case of two recent immigrants from the Near and Middle East

Parsis and Jews are the Near East and Iran-related communities that migrated to South Asia during the historical period, which have been studied genetically.

2.7.1. Parsis

The South Asian Parsi community belongs to the Zoroastrianism religion that follows the Prophet Zarathustra (Zoroaster) and thrived in Persia (modern Iran) from 600 BC to 650 AD (Mirza, 1974; Hinnells, 1994, 2007). The dwindling number of Zoroastrians count around 140 thousand worldwide; majority of them currently reside in India, and a tiny fraction lives in Pakistan, while the original homeland – modern Iran – has a considerably smaller share of Zoroastrian inhabitants relative to India (Shroff and Castro, 2011; UNESCOParzor, 2019). The declining population size of the Parsis is linked to their relatively low birth rates, followed by strict endogamy practices within the community members (Shroff and Castro, 2011; UNESCOParzor, 2019). In the attempt to protect themselves from the Islamic conquest of Arabia during the seventh century, they arrived in the modern Indian state of Gujarat and got their name 'Parsi', referring to their root from Persia that local Indians called Paras or Fars (Sekar, 1948; Mirza, 1974; Roy et al., 2001; Hinnells, 2007). Linguistically, Parsis of the Indian subcontinent are a group of Indo-European speakers. The South Asian Parsi community, though assimilated well with the local people linguistically and economically by embracing the local Gujarati and Sindhi languages and for their significant contribution to the South Asian economy, yet maintained ethnic uniqueness by marrying strictly within their community (Mirza, 1974; Karkal, 1975, 1984; Palsetia, 2001; Hinnells, 2007; Jonnalagadda et al., 2011; Shroff and Castro, 2011).

Early genetic studies of Parsis mostly used low-resolution markers, analyzing individuals of the Parsi community mainly from Pakistan (Undevia, 1973; Qamar et al., 2002; Quintana-Murci et al., 2004; Mohyuddin and Mehdi, 2005; Rosenberg et al., 2006). Indian Parsis were underrepresented in these studies despite their predominant contribution (~98%) to the Parsi population worldwide. They detected that the genetic variation of Parsis lies midway through the contemporary South Asian and West Eurasian genetic variations (Qamar et al., 2002; Quintana-Murci et al., 2004; Mohyuddin and Mehdi, 2005; Rosenberg et al., 2006). Parsis's male-biased assimilation pattern with local South Asian females was also observed (Quintana-Murci et al., 2004). The study analyzed maternal markers and found that 60% of the maternal lineages among Pakistani Parsis are South Asia specific (Quintana-Murci et al., 2004). In contrast, paternal male lineages prevalent in Parsis are almost entirely related to Iran (Qamar et al., 2002).

Parsis possibly endured broader genetic isolation to retain their root heritage, which resulted in higher recessive disease carriers among the Parsi community. Therefore, a genetic survey of the Parsi community applying high-resolution markers and large sample sizes from their different habitats is a highly sought

interest. Comparing the underrepresent Parsi individuals living in India, including their ancient remains, to available modern and ancient specimens across Eurasia is especially significant in this regard. To address this, two recent papers (López et al., 2017; Ref II) analyzed biparental and uniparental genetic data of Parsi individuals from South Asia and Iran, comparing them with modern individuals worldwide. In another study (Ref III), we analyzed the Indian Parsi individuals compared to the available aDNA data from West Eurasia and South Asia. The details of our study (Ref III/IV) are discussed in the results & discussion section of the thesis. Lopez et al. (López et al., 2017), in a similar outcome to our study (Ref III), observed a higher genetic similarity of South Asian Parsis to modern Iranians than with any other sampled modern group and also indicated continuing high endogamy among Parsis. The studies also traced the admixture between Iranian and Indian Zoroastrians between 690–1390 CE.

2.7.2. Jews

The worldwide scattered Jewish populations are shown to be a genetic mosaic of peoples of autosomally, patrilineally and matrilineally different origins, who historically follow elements of Judaistic tradition (Atzmon et al., 2010; Behar et al., 2010, 2013; Yardumian and Schurr, 2019). Jewish diaspora immigrated to the Indian subcontinent in three known episodes of migrations at least, setting up the stage for the rising of three distinct Jewish communities at three different coastal areas of India; the Cochin Jews in Kerala port city Cochin, the Bene Israel Jews in Mumbai, and the Baghdadi Jews in Kolkata (Katz, 2000; Israel, 2002). The community of Cochin Jews now comprises another Jews group, the fourth one – the Pardesi Jews, who perhaps migrated to India from Spain and Portugal during 15–16th centuries, followed by their assimilation into Cochin Jews (Katz, 2000). The Cochin Jews were likely the early migrants, but with a smaller number of individuals than the next incoming Bene Israels, who were tagged as the largest Jewish community of India (Slapak and Yiśraʾel, 1995; Ehrlich, 2009; Roland, 2018). Several well-known narratives stated a range of time, from pre and early historical to early middle period, for the entrance of Jewish communities into India, but no clear cut clue was established regarding their origin and immigration, mainly because of the lack of written records and engravings (Slapak and Yiśraʾel, 1995; Katz, 2000; Israel, 2002; Fernandes, 2011). The exception is the Baghdadi Jews, who migrated to the Kolkata port of East India recently, during the 19th Century British rule (Ṿail, 2002).

Despite the inadequate specimens and low resolution, early genetic studies explored the Cochin Jews and Bene Israelis. They observed that Indian Jewish communities share a varying degree of genes with both the neighboring Indian groups and Middle Easterners, thereby indicating a pronounced admixture of Jews with local Indians (Cohen et al., 1980; Metspalu et al., 2004; Behar et al., 2008a, 2010; Rootsi et al., 2013). Classical markers linked the Cochin Jews to Yemenite Jews (Cohen et al., 1980), while chrY connected the Bene Israeli community to the Levant (Rootsi et al., 2013), and mtDNA studies on Indian Jews detected

predominant South Asia related maternal lineages among Indian Jews (Behar et al., 2008a, 2010). Recently, a few studies analyzed Jews of Cochin and Bene Israeli communities using larger sample sizes and high-resolution markers and obtained more detailed information about their origin and migration, including the admixture dates (Chaubey et al., 2016; Waldman et al., 2016a, 2016b). However, the studies above commonly found a greater genetic sharing of both the communities with local Indians and a minor but considerable contribution from Jews-related ancestry, which plausibly reflects the consequence of predominantly female-biased mixing from the Indian side. A few studies sampled individuals from the Indian Jewish community, living in Israel now, showing that mostly Yemenite, Sephardi, and the Middle East-related Jews admixed with South Asians ~400–700 yrs ago and ~700–1100 yrs ago respectively, forming the Cochin and Bene Israeli Jews (Waldman et al., 2016a, 2016b). Nevertheless, another study that mostly used Indian Jewish individuals living in India perceived the admixture events ~1500 yrs ago and ~1000 yrs ago responsible for forming the Cochin and Bene Israeli Jews, respectively (Chaubey et al., 2016). Later migration of Bene Israeli Jews (Mumbai Jews) to India is also reflected in their higher genetic sharing with ancestry, related to the Middle East peoples, than that of the Cochin (Kerala) Jews.

When compared to modern populations, genetic patterns displayed by different Jewish communities make them an interesting case to investigate whether such patterns also remain when they are examined with available aDNA data from West Eurasian individuals. Besides, seeing the prevailing local Indian-like ancestries in both the Kerala and Mumbai Jews, modeling both the Jewish communities with the same set of ancient sources that were used to model Indian populations, can be another interesting aspect that is further discussed in the results and discussion section (Ref III) of this thesis.

2.8. Particular case of linguistically diverse Tharu tribe from Tarai region

Naturally diverse Tarai area, found in the northern section of the Indian subcontinent, connects North India and southern Nepal. Many ethnically different populations with diverse languages and cultures live in the Tarai region (Guneratne, 2002). Among them, one of the most prominent groups, the Tharu people, is plausibly the most linguistically diverse and spread throughout the northern Indian border areas (Uttarakhand, Uttar Pradesh, Bihar) and South Nepal. Tharu ethnic name is comprised of many culturally diverse groups, speaking different dialects and practicing inbreeding. Different groups of Tharu people use different kinds of Indo-Aryan languages. They are recognized as scheduled tribes in the Indian states of Uttarakhand and Uttar Pradesh, while those living in Nepal are official nationals of Nepal. Though different Indo-Aryan dialects used by Tharu people are grouped as *Tharuvani,* none is recorded, leaving no single existing language of the *Tharuvani* group, thus complicating the reality of the Tharu language. In general, Tharu people interact with their IE-speaking neighbors in their versions of native IE varieties, such as Bhojpuri, Awadhi, Urdu, and Maithili.

On the other hand, Tharu people also use a different sort of language that appears to be close to TB when they speak to their inner circle people (Van Driem, 2001). Tharu people mostly remained isolated in their own neighborhoods for hundreds of years, developing a distinct culture unaffected by nearby people from Nepal highlands and India (Guneratne, 2002) and claiming to be the folks of the jungle. Tharu are spiritually connected to nature and pray to gods living in forests before entering the forest (McLean, 1999). Vocal stories and folklores blur the origin of the Tharu people; some of Tharu groups claim to be related to Rajput immigrants from the Thar Desert, whereas other Tharu groups, living more eastward, claim their ancestry in Sakya and Kolia populations of Kaplivastu. However, several studies (Terrenato et al., 1988; Modiano et al., 1991; Passarino et al., 1992) have suggested that Tharu people, despite their existing diverse speaking pattern, are related to a common ethnic group with long term natural settlement in the Tarai region (Chhetri, 2005).

Following a WHO-supported program to eradicate malaria in the Tarai region, genetic studies helped to realize a shared connection among different ethnic groups of Tharu. They are way less prone to malaria than their neighboring non-Tharu populations. Specifically, the alpha thalassemia genetic variant that protects against malaria is predominant among all Tharu peoples (Terrenato et al., 1988; Modiano et al., 1991). The finding suggests that the genetic variant, enhancing the protection against malaria among Tharu groups, must have evolved through natural selection due to their continued contact with the disease-favoring environment full of swamps and forests, endorsing the likely origin of Tharu people in vast Tarai jungles. The use of TB-related words by linguistically heterogeneous Tharu groups when they converse within close groups, agree with the prevalence of East Asian-related genetic component among Tharu people, traced by uniparental and anthropometric studies (Passarino et al., 1992, 1993; Koirala et al., 2012). Another study that analyzed the mtDNA and chrY from a couple of Tharu-populated locations in Nepal suggested that the ancestry of genetically diverse Tharu groups contains West Eurasian and Indian-specific genetic contributions, in addition to the East Asian one (Fornarino et al., 2009). Most of the Tharu maternal and paternal lineages share a deep ancestry with Indians; remarkably, the Tharu people harbor local paternal ancestries present in South Indian tribes. Whereas mtDNA reveals more diversity among Tharu groups, together with their genetic sharing and ancient connection with populations from Malaysia, Japan, and Andaman Islands (Fornarino et al., 2009).

The interesting location, blurry origin, and linguistic heterogeneity make Tharu an exciting case population for a detailed study by taking advantage of modern high-resolution experimental approaches. In particular, previous genetic studies have analyzed only mtDNA and chrY data, which are more susceptible to random genetic drift than genome-wide data. Though genome-wide data for the Tharu people have been generated earlier (Reich et al., 2009; Metspalu et al., 2011), questions like reconstruction of their genetic history and how they admixed with different ancestral source populations have remained obscure, which we tried to address in Ref IV of the thesis.

3. OBJECTIVES OF THE STUDY

The broader aim of the current thesis was to study the demographic dynamics of contemporary South Asian populations, more concisely, the existent settlers of the modern Indus valley region of Northwest India, a few historical immigrants from the Near and Middle East, and a linguistically diverse ethnic group of the Tarai region. In general, the present study looked for genetic portraits of these populations relative to neighboring population groups. Besides, it took advantage of emerging aDNA studies that allowed the possibility to explore the relative affinity of these contemporary peoples to the ancient inhabitants of West Eurasia and South Asia. Specifically, the objectives of the studies were as follows;

1. Ref I), to learn the genetic architecture of populations from underrepresented Northwest India, and present-day genetic variation in the historical Indus Valley region.
   - To sample blood or saliva from recruited healthy donors who volunteered to participate in the study, and extract DNA.
   - To generate uniparental and genome-wide SNP data from four populations of Northwest India, and analyze the uniparental genetic patterns.
   - To analyze and compare the new genome-wide data with that of available modern and ancient genomic data of neighboring regions and West Eurasia, and get insight into the understated genetic structure of the present-day Indus Valley region.
2. (Ref II & III), to reveal the admixture dynamics of a few historical Near and Middle Eastern migrants to South Asia in the last two millennia.
   - To generate the genome-wide array data and high-resolution uniparental data from Parsi populations in South Asia, in addition to mtDNA genotyping data from ancient Parsis.
   - To characterize the population structure of current Parsi peoples living in South Asia, underlying newly generated autosomal and uniparental data compared to previously published global population genomic data.
   - To reconstruct the admixture and integration pattern of Parsis and Jewish populations by analyzing the available genomic data and relating them to that of ancient individuals from the neighboring regions.
3. (Ref IV), to decode the distinct strains of genetic ancestry in Tharu tribe from the northern Tarai region of the Indian subcontinent.
   - Reconstruction of genetic history of Tharu group by estimating the ancestry proportion of different ancestry components, using earlier published autosomal data of Tharu and comparing them with Eurasian population genetic data.
   - To evaluate different Tharu people's intra and inter-group relationships with neighboring populations, comparing both the uniparental and the autosomal genetic data.

4. MATERIALS AND METHODS

The subjects investigated in four individual publications of the current study are described in respective research articles and their supporting materials.

The DNA samples were obtained from blood or saliva of unrelated healthy adults, who volunteered to participate in the study after receiving informed consent, in accordance with the guidelines approved by the Ethical Committees of the institutions involved.

The methods applied in the present study, including the experimental and the computational, are described in detail in the corresponding publications and supporting materials thereof.

5. RESULTS AND DISCUSSION

The results of four articles published in peer-reviewed journals are discussed and presented here as a summary of the novel outcomes of the current study. This section aims to offer an outline of the main results of the published articles; however, it does not recollect the contents of the articles word by word; additional details can be found in the respective publications and their supporting materials.

5.1. Genetic variability among modern Indus Valley populations suggest a population-specific demographic history and ancient links (Ref I)

In this study, we generated and analyzed 45 new genome-wide genotype data (Supplemental data – Table S1 in Ref I) and uniparental markers data for ~200 individuals of four populations from the eastern Indus Valley region of Northwest India (NWI), which otherwise was poorly covered so far from diploid genome aspect. We pooled this data with previously published 20 genome-wide data from the Punjabi Khatri population (Basu et al., 2016) of NWI and compared them with that of hitherto published modern and ancient samples across the globe.

5.1.1. Genetic ancestry composition of populations from east and west Indus basin

We analyzed our pool of 65 genome-wide data from NWI, comparing them with available modern samples (Li et al., 2008; Behar et al., 2010, 2013; International HapMap Consortium, 2010; Chaubey et al., 2011b; Metspalu et al., 2011; Rasmussen et al., 2011; Yunusbayev et al., 2012, 2015; Cristofaro et al., 2013; Fedorova et al., 2013; Raghavan et al., 2014; Kushniarevich et al., 2015; Basu et al., 2016) (Supplemental data – Table S1 in Ref I). Our allele frequency-based ADMIXTURE, PCA, and F_{ST} analyses (Figure 1B, Supplemental data – Figure S2, Table S5 in Ref I) found that, in general, NWI individuals at the eastern Indus basin are genetically close to their neighbors living on either side of them. In short, NWI populations are a genetic intermediate of population groups that reside in the western Indus region of Pakistan and eastward lower region of the Gangetic Plain. These results were further endorsed by our haplotype-based identity by descent (IBD) (Browning and Browning, 2007, 2013) and CHROMOPAINTER (Lawson et al., 2012) **(Figure 10;** Supplemental data – Figure S13 in Ref I) analyses. This apparent closeness of NWI to Pakistanis (west of Indus) and North Indian Indo-Europeans (NI_IE) living in lower Gangetic Plain, contrast to former uniparental data-based studies (Quintana-Murci et al., 2004; McElreavey and Quintana-Murci, 2005). Comprehensively, such a pattern might agree with the

archaeological suggestions of high mobility of Indus Valley people, including their likely migration toward the eastward Gangetic Plain after the Indus Valley Civilization collapsed (Misra, 2001b; Tripathi et al., 2004; Gupta et al., 2006; Madella and Fuller, 2006; Brooke, 2014).

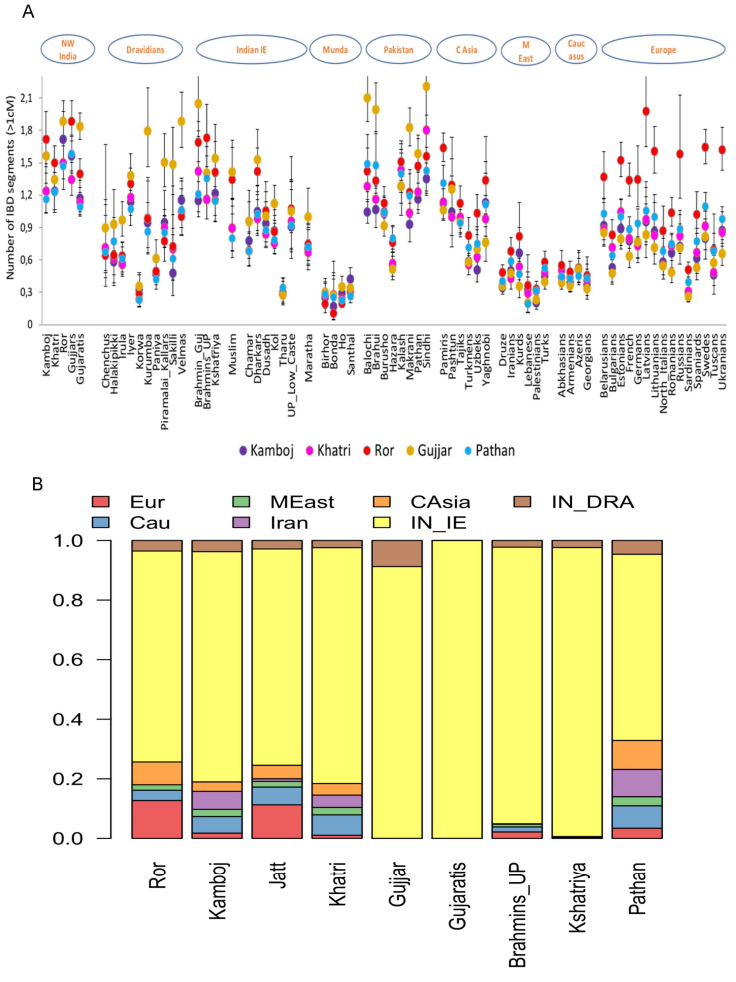

Figure 10. IBD Sharing and Ancestry Profile of modern Indus Valley Populations.

(A) Average number of IBD segments per pair of individuals for each Northwest Indian and Pathan ethnic group from Pakistan. Error bars represent 95% confidence intervals. **(B)** Population-based ancestry estimates for PNWI and neighboring groups from the North Indian Gangetic Plain were inferred by CHROMOPAINTER via an NNLS-based analysis. NWI and Pakistani populations (PNWI) were excluded from the donor groups. Abbreviations are as follows: IN_IE, Indian Indo-Europeans; Pak, Pakistan; Eur, Europe; M East, Middle East; Cau, Caucasus; and IN_DRA, Indian Dravidians. Reprinted from Figure 2 of Ref I, with permission from "Elsevier" under the Creative Commons CC-BY-NC-ND license.

We also detected a considerable genetic variability among the contemporary populations of the Indus region **(Figure 10**; Supplemental data – Figure S13 in Ref I) that fits well with the isolation by distance model. Ancestry profiling-NNLS results later confirmed the same pattern **(Figure 10B**; Supplemental data – Figure S13 in Ref I). A few Northwest Indian groups like Ror (and Jat) are genetically more attracted to modern Europeans and Central Asians **(Figure 10;** Figure 1B–D, Supplemental data – Figure S5/S13 in Ref I), compared to other Northwest Indian counterparts. This higher European/Central Asian affinity, together with the high relatedness of Ror (and Jat) to Pathan and Kalash groups in PCA, $F_{ST,}$ and *D*-statistics results (Supplemental data – Figure S2, Table S5/S7/S8 in Ref I), point a joint Central Asian/European connection of these groups. However, the ALDER result did not find any evidence of recent admixture, of these or any other groups, with Ror for the last ~1500 years (Supplemental data – Table S18 in Ref I). In contrast, other NWI groups, such as Gujjar, are genetically close to local Indians. Collectively, these outcomes signify unrelated ancestral components and long-term population structure in the Indus region, which is later supported by their distinct allele sharing pattern with ancient individual sources (Supplemental data – Figure S6-S9, Table S9/S10 in Ref I).

This unique visible pattern of genetic heterogeneity among modern Indus populations, rarely seen among other regional Indian populations, probably connects them to the ancient Indus Valley people, who believably were quite heterogeneous. We also used Runs of Homozygosity (RoH) (Supplemental data – Figure S12 in Ref I) and clarified that the close affinity of Ror to the Europeans is not contributed by inbreeding.

5.1.2. Modern Indus Valley populations in terms of ancient West Eurasian ancestries

In addition to the previously published global modern data, we also compared our new autosomal data from NWI to that of available West Eurasian ancient DNA data (Supplemental data Table S1, S2 in Ref I) (Allentoft et al., 2015; Haak et al., 2015; Mathieson et al., 2015; Lazaridis et al., 2016; Damgaard et al., 2018; Narasimhan et al., 2018). We analyzed the merged data applying f_3, *D* statistics, and *qpAdm* (Figure 3, Supplemental data – Figure S6-S9, Table S9-S16 in Ref I). Our results agree with previous claims that the extensively dispersed ancient West Eurasian component among South Asian populations was contributed by a combination of ancient sources (Broushaki et al., 2016; Lazaridis et al., 2016; Damgaard et al., 2018; Narasimhan et al., 2018). These ancient sources were related, either distantly (Iran_N and Steppe_EMBA) or closely (Namazga_CA/ Indus_Periphery and Steppe_MLBA), to South Asians in space and time. We detected that relative to contemporary South Asian groups, modern Indus Valley people showed a higher affinity with ancient Mesolithic to Iron Age West Eurasian individuals (Supplemental data Figure S7-S9, Table S9/S10 in Ref I). This trend signifies the geographic location of the Indus Valley that have been

used as a gallery to South Asia for all sort of influx from western regions (Majumder and Basu, 2015).

The modern Indus Valley people display greater genetic closeness to the Neolithic farmers of Iran and Anatolia (Supplemental data – Figure S6C/S7B/S8B, Table S9/S10 in Ref I) than other Indian counterparts. They also show a higher Middle Eastern component (Figure 1B in Ref I) and prevalence of chrY hg J2-M172 (Supplemental data – Table S4 in Ref I). Y chromosomal hg J2-M172 has been linked to the substantial incursion of Middle Eastern male lineage through the northwest route (Singh et al., 2016). Taken together, these patterns lend credence to former archaeological suggestions of a potential impact from the Zagros or Levant region over the development of agriculture and settled life in South Asia (Jarrige, 1995; Gangal et al., 2010).

5.1.3. Unique genetic structure of Ror population

Importantly, our study reports Ror as the only South Asian group with significant allele sharing to Neolithic farmers of Anatolia than that of Iran (Figure S6B/S8B, Table S9 in Ref I), including their higher genetic intimacy to Caucasus hunter-gatherers (CHG) relative to other subcontinental populations. These observations probably simplify interpretation of the previously reported extra prevalence of the light skin color variant (*SLC24A5* allele *rs1426654*) in the Ror population (Gallego Romero et al., 2012; Mallick et al., 2013), since this allele has been highly frequent in both the ancient individuals (Jones et al., 2015; Mathieson et al., 2015). Additionally, we found that Ror (and Jat) also displayed higher genetic closeness to EHG, Bronze Age Steppe herders (Steppe_EMBA and Steppe_MLBA), and modern Europeans than that among other South Asian groups. This pattern also syncs well with the earlier reported high prevalence of lactase persistence gene variant (*LCT-13910T*) in Ror (and Jat) (Gallego Romero et al., 2012; Mallick et al., 2013). Overall, this suggests two potential causal scenarios; either a recent Europe-mediated gene flow or an interaction with ancient West Eurasian individuals. Strikingly, we observed that higher IBD sharing between Ror and Europeans positively correlates to the increasing Steppe ancestry in Europeans (Supplemental data – Figure S15 in Ref I). Older admixture date and higher affinity to Pathan and Central Asians, coupled with the declining genetic affinity of ancient individuals of Steppe, Anatolia_N, CHG, and EHG from modern Centrals Asians to Ror, suggest a Central Asian/Steppe connection for the higher modern European and ancient West Eurasian related affinity of Ror. Furthermore, the highest ancestry proportions of both the Early and Middle Bronze Age Steppe herders in Ror (Figure 3, Supplemental data – Table S11/S12 in Ref I) indicate multiple Steppe-like inflows in the Ror. The Steppe ancestry is the likely key for the greater genetic intimacy of the Ror group to the pre- and early historical ancient South Asians from Swat Valley (Figure S6A-B/S7/S8A/S8D/S9, Table S9/S16 in Ref I).

Furthermore, we compared the results of uniparental data to that of autosomal data for the Ror group. We found that the apparent West Eurasian ancestry connection of Ror, displayed by genome-wide SNP data, is not braced by maternal mtDNA data. In comparison, paternal Y chromosome data show a higher frequency of West Eurasia-related haplogroups (Supplemental data – Table S3/S4 in Ref I) henceforth, suggesting a plausible male-dominated admixing from the West to Ror.

5.1.3.1. Alternate ANI representative

Ror and Kalash groups display a distinctly close genetic relation and equally higher proportions of ANI component compared to other South Asians. That, alongside the advantage of Ror in having diverse Steppe ancestries, points out the Ror as a potential alternative to Kalash in the representation of the ANI component. Therefore, Ror can plausibly be used as an ANI proxy in the demographic modeling of modern Indians than the earlier proposed Kalash, which is extensively privately drifted (Ayub et al., 2015b).

To summarize, our study revealed a substantial higher genetic variability among present-day populations living in the Indus Valley region, which possibly is linked to the long-term population structure from the prehistorical period. We noticed a distinct and positively correlated genetic affinity of Ror to modern Europeans and ancient Steppe individuals. Additionally, we recognized a population-specific demographic process in the culturally distinct groups of Northwest India. Interestingly, the study reported Ror as the so far only Indian group that harbors a higher Anatolian farmer ancestry than the Iranian farmers.

5.1.4. Recent updates from the first aDNA of Indus Valley *compared to* our study

The first successfully sequenced single Bronze Age aDNA from Rakhigarhi, the Indian side settlement of the Indus Valley civilization, has modified the anticipated population history of prehistorical South Asia (Shinde et al., 2019). The study found that the Bronze Age IVC individual comprised the genetic ancestries related to the ancient Iran and Southeast Asian hunter-gatherers, the two groups that contributed to modern South Asians' genetic ancestries, too. Thus, the study settles the Bronze Age IVC individual as the probable primary source of ancestry to all contemporary South Asians. However, the study suggested that ancient Iran-related ancestry, observed in the IVC individual, was related to an ancestor group that survived before the separation between ancient Iranian hunter-gatherers and agro-pastoralists. Thus, the Iranian ancestry of South Asians comes from an ancient Iranian group that was perhaps living before the farming in the Fertile Crescent (FC) started. Therefore, the new findings attempted to amend the earlier beliefs that FC-related agriculturalists likely initiated agriculture in South Asia

(Jarrige, 1995; Gangal et al., 2010). The study claimed that farming in the subcontinent plausibly developed either independently or as a consequence of the exchange of skills through cultural contacts. Another notable update to the population history of South Asians comes through the absence of Steppe ancestry in this Bronze Age IVC individual.

Interestingly, the new aDNA of Indus Valley is genetically identical to the Indus_Periphery group, which we have also used in our study's modelings (Ref I). The new study suggests that contemporary Indians comprise ancestry from Steppe people topped on the Indus Valley individual ancestry, which is similar to what we have done. Nevertheless, since the new study claimed an altered source for Iranian ancestry, the foremost model we used in our study (Steppe + Iranian Neolithic farmers + Onge) might not have been that appropriate. Additionally, the investigation of IVC aDNA by Shinde et al. (Shinde et al., 2019) denies the Fertile Crescent connection to the development of agriculture in South Asia that we discussed in our results. However, one must bear in mind that the aDNA data available on this topic is thus far absolutely minimal for any final generalizations.

5.2. Genetic outline of historical migrants to South Asia: The case of Parsi and Jewish populations (Ref II/ Ref III)

The majority of the South Asian populace have been following endogamy and restricted marriage practices for several thousands of years (Chaubey et al., 2007; Basu et al., 2016b; Nakatsuka et al., 2017; Narasimhan et al., 2019). Such an old existing endogamous social structure probably must have influenced the commixture and absorption drive of foreign migrant groups robustly (Behar et al., 2010; Shah et al., 2011; Chaubey et al., 2016; López et al., 2017) (Ref II/III). To understand this impact, we used high-resolution data from two populations with the record of historical migrations to South Asia. We generated 43 new genome-wide genotyping array and high-resolution uniparental genotyping data from 174 Indian and Pakistani Parsis, in addition to 21 ancient Parsi mtDNA genotyped data from Gujarat. Later, we compared them with worldwide population genetic data from modern and ancient individuals. In the case of Jews, we used previously published Jews genome-wide data (Atzmon et al., 2010; Behar et al., 2010) and analyzed them in the context of available modern and ancient data published previously.

5.2.1. Population structure and demographic history of Parsis

Previous efforts (Undevia, 1973; Qamar et al., 2002; Quintana-Murci et al., 2004; Mohyuddin and Mehdi, 2005; Rosenberg et al., 2006) to study the genetic history of the Parsi population had limited analysis power. It was mainly due to two factors: a) the use of low-resolution uniparental markers and b) lacking the representation of Parsis from India, where now the majority of Parsis (>98%) live. Here, we analyzed new autosomal SNP data from both the Indian and Pakistani Parsis, with

that of global populations, to answer questions like; how much genetically related both Parsis groups are, mutually, and to the modern Iranians (considering their origin). Besides, we looked at the extent of their blood kinship and their genetic relation to available ancient West Eurasians. We also tried to get insight into how efficiently Parsis integrated with local females and sex-dependent demographic processes, if any, by exploring mtDNA and Y chromosome genotyping.

5.2.1.1. Genome-wide ancestry of Indian and Pakistani Parsis

In the context of modern populations

Our PCA, ADMIXTURE, and F_{ST} results (**Figure 11A;** Fig. 2/3, Additional files – Fig. S1/S2a, Table S3/S4 in Ref II) disclose that the Indian and Pakistani Parsi groups are genetically closest to each other than any of their South Asian neighbors. Their closeness to modern Iranians was boosted by their intermediate positioning between Pakistani and Iranian groups when the population-wise mean of eigenvalues was plotted in PCA (Fig. 2 in Ref II). Both results, collectively, direct a common genetic heritage of Indian and Pakistani Parsis from their presumed birthplace in Iran. This might also be a valid explanation for their apparent and significantly higher affinity to West Eurasian populations than the South Asian ones in F_{ST} result, which was re-captured by ADMIXTURE analysis as well. Altogether, it specifies a likely major West Eurasians (Iranians) ancestry contribution to the Parsis genome and a lesser South Asian contribution. Which, bolstered by further results of f_4 ancestry estimates, outgroup f_3, and *TreeMix* (Additional files – Fig. S2b/S4-S7 in Ref II), backs the historical understanding that Parsis share most recent common ancestry with Iranians.

We observed a highly significant level of gene flow among Indian and Pakistani Parsis, akin to their immediate South Asian neighbors (Gujarati and Sindhi) has with each other, in *D* statistics (Additional file – Table 2 in Ref II). However, South Asian Gujarati and Sindhi groups also have higher gene flow with Parsi groups relative to contemporary Iranians. They also donate considerably higher chunks to the Parsi genomes than they have received back. These results, in addition to the lower chunk counts sharing of both groups with Parsis than between any South Asian groups in fineSTRUCTURE (Additional file – Fig. S9 in Ref II), mostly articulate a slight and one-way gene flow from South Asians to Parsis.

Furthermore, we used ALDER and MALDER to estimate, when Parsis admixed with their neighboring South Asian Gujarati or Sindhi groups and whether they went through multiple admixture events. We found the evidence of a single admixture event that took place ~1200 years, thereby, agreeing well with the historical records of Parsi migration to South Asia (Additional file – Table 3 in Ref II). We also detected a unique demographic history of both the Parsi groups, characterized by a greater number of long runs of homozygosity (RoH) segments in Parsis (Additional file – Fig. S10 in Ref II) than their current neighbors and alleged parental population. This indicates a possible sign of lesser population mass and higher level of inbreeding or consanguinity in Parsis.

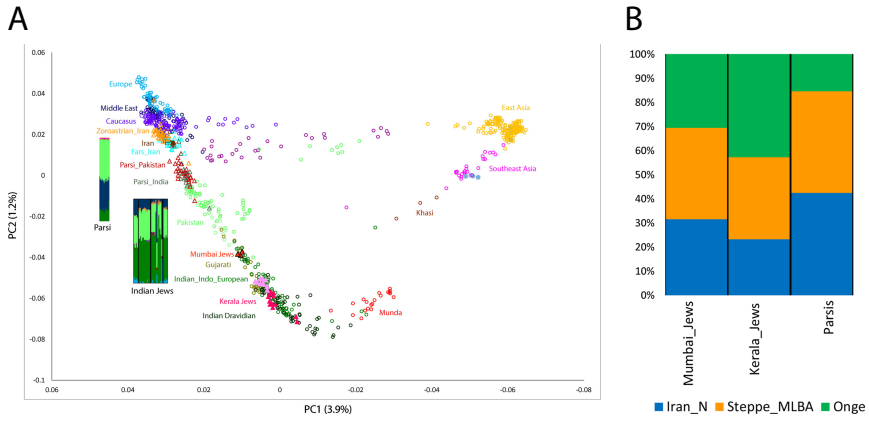

Figure 11. Population structure and ancient ancestry profile of Parsi and Jews.

(A) PCA of Eurasian populations showing the placement of Indian Jewish and Parsi groups along with the South Asian cline. It is noteworthy that they were attracted towards West Eurasian populations, suggesting a higher level of ancestry related to West Eurasians complying with their neighboring local South Asian populations. The bar plot shows ADMIXTURE proportions of both of the studied populations. **(B)** Plot of ancestry proportions modeled in the studied populations. Adapted from Figure 1 and 5 from Ref III, with permission from "Springer Nature".

Later, we used PBS analysis to examine the effect of genetic drift on functional genetic variation in Parsis after the occurrence of admixture events. We found a group of associated SNPs in the HLA region and a CD86 related missense SNP (rs1129055) with a more frequent ancestral G allele in Parsis than Iranians, South Asians, and East Asians (Additional files – Fig. S11, Table S8 in Ref II). The SNP (rs1129055) is linked with sepsis prompted by pneumonia, and its G variant has been related to the reduced threat of active brucellosis in Iranians (Eskandari-Nasab et al., 2014). It has a potential role in minimizing the B lymphocytes infection from the Epstein-Barr virus (GTEx Consortium, 2015).

In the context of ancient individuals

Parsis display higher genetic affinity to modern Iranians (discussed above), accompanied by significantly higher Near East blue ancestry than South Asians. Remarkably, they also have a lesser European light blue ancestry and a noticeable higher South Asian dark green ancestry compared to modern Iranians (Fig. 3, Additional files – Fig. S2a, Table S4 in Ref II). This prompted us to apply f_3 and D statistics to analyze the combined genomic data of Parsis and ancient Neolithic Iranian farmers since early Neolithic farmers from Zagros have reportedly been found genetically closer to Zoroastrians of Iran (Broushaki et al., 2016). The resulting outcomes (Table 2, Additional Files – Figure S3, Tables S5/S6 in Ref II) suggest a considerably higher genetic affinity between Parsis and Neolithic

farmers when compared to the modern people of Iran. Modern Iranians likely originated from an ancestral lineage that was a product of interbreeding between Neolithic Iranians and Arabian Peninsula populations whereas Parsis did not. Accordingly, our results support a likelihood that extra light blue ancestry present in modern Iranians (Fig. 3 in Ref II) was introduced later, as a result of the mixing with Arabic peninsular groups during the Islamic conquest after the Parsis had left Iran (Grugni et al., 2012; Derenko et al., 2013; Haber et al., 2013).

Our study also reveals a higher affinity of Parsis to most ancient West Eurasian individuals (Table 1 in Ref III), similar to the pattern they showed with modern West Eurasians, thereby confirming a likely West Eurasian related origin for Parsis. Additionally, Parsis harbor the highest West Eurasian ancestry proportion (~85%) in South Asia and the lowest South Asian-related ancestry proportion (~15%), when we applied *qpAdm* (**Figure 11B**) using Neolithic Iranians, Steppe_MLBA, and Onge as three ancient sources to model modern South Asians. Somewhat expectantly, the highest West Eurasian ancestry proportion observed in Parsis comprises equal amounts of both the Neolithic Iranian farmers and Steppe people, unlike the contemporary South Asians (Ref I). Altogether, these results again endorse the West Eurasian link of Parsis and their subsequent minor intermixing with the contemporary South Asian neighbors as suggested earlier. However, the higher affinity of Parsis to East Asian related Ust'-Ishim than modern Iranians is expected, considering the surplus South Asian ancestry in Parsis compared to Iranians.

5.2.1.2. Maternal and paternal lineage of Parsis

To understand how the Parsis are affected by male or female genetic components of South Asia and Iran, we examined mtDNA and chrY genotyping markers for a higher number of Indian and Pakistani Parsis (Fig. 4, Additional files – Fig. S12, Table S9/S10 in Ref II). We, too, compared our results to the mtDNA of 21 ancient Parsi individuals (Additional file – S11 in Ref II) from Sanjan (Jonnalagadda et al., 2011), identifying the founding maternal lineages and also lineages that integrated later. Our study found ancient Parsis having 48% maternal lineages related to South Asians, which indicate either these haplogroups have been carried by the Zoroastrian migrants from Fars (Iran), owing to their considerable presence (~10%) in contemporary Persians reported earlier (Derenko et al., 2013). Alternatively, it might also reflect an earlier settlement of Parsis, followed by their mixing with the local South Asian females. In sum, we observed a greater prevalence of Middle-Eastern haplogroups in the chrY pool of Parsis (Additional file – S12, Table S10 in Ref II) compared to their mtDNA pool.

5.2.2. Jews in the context of ancient West Eurasian individuals

The Indian Jews, apparently the old offsprings of the exclusive Jews diaspora of the world (Ehrlich, 2009), are categorized into three groups based on their inhabitation: Kerala Jews living in Cochin, Bene Israel Jews of Mumbai, and Baghdadi Jews residing in Kolkata (Katz, 2000). Among all Jewish diaspora, the Indian Jews have been one of the least studied such ethnic groups (Slapak and Yiśra'el, 1995; Katz, 2000; Israel, 2002). However, three recent studies attempted to survey their genome-wide data more thoroughly (Chaubey et al., 2016; Waldman et al., 2016a, 2016b). These studies followed a different sampling approach wherein, Chaubey et al. used samples from Indian Jews living in India. The other two studies by Waldman et al. sampled Indian Jews individuals who had returned to Israel.

Collectively, these studies established a distinctive genetic sketch of Indian Jews. Though closely related to local South Asians yet, Indian Jews have heterogeneous genetic character agreeing to the earlier claimed idea that Jews populations were established by the union of peoples with different ancestries (Yardumian and Schurr, 2019). Mumbai Jews and Kerala Jews show a sort of heterogeneity, corroborated by the extra Middle East-related genetic component in Mumbai Jews than in Kerala Jews (**Figure 11**; Figure 2 in Ref III). This can be explained through isolation by distance model and even fits well with the record of the earlier settlement of Kerala Jews in India than the Mumbai Jews (Chaubey et al., 2016). The study had also observed a higher genetic sharing of Indian Jews to the populations living in the Middle East and Jews, relative to their neighboring populations of India, reflecting their deep connection to the Near East and historical Jewish populations. This relationship explains the impact of the admixture event between 1100–1600 CE when a population ancestral to Indians intermixed with the ancestral group of Jews from the Near East (Chaubey et al., 2016). Uniparental markers had revealed the higher prevalence of South Asian related maternal lineages coupled with pervasive Near/Middle East-specific lineages, therefore, suggesting a male-mediated admixture from Jewish migrants in South Asia (Chaubey et al., 2016).

However, considering the growing number of ancient samples, it is imperative to understand the genetic relation of Indian Jews in the context of ancient individuals from West Eurasia and South Asia (Allentoft et al., 2015; Haak et al., 2015; Mathieson et al., 2015; Lazaridis et al., 2016; Damgaard et al., 2018; Mittnik et al., 2018; Narasimhan et al., 2018). We studied the earlier published genome-wide SNP data of Indian Jews (Atzmon et al., 2010; Behar et al., 2010), applying f_3, *D* statistics, and *qpAdm* to fulfill this objective. We observed a substantially higher affinity between Mumbai Jews and ancient Steppe people in both the f_3 and *D* statistics analyses compared to South Asian Dravidians (Figure 4, Table 1 in Ref III). Subsequently, the result was also validated by *qpAdm* in Mumbai Jews having higher ancient Steppe-like ancestry proportions than seen in Kerala Jews (**Figure 11B**; Table 2 in Ref III). In contrast, Kerala Jews are genetically related to ancient individuals of West Eurasia (including Steppe) and

South Asia as equally as their geographical neighbors (Dravidians) are. This pattern was even reflected by lesser ancient West Eurasian and higher South Asian genetic component in Kerala Jews than in Mumbai Jews (**Figure 11B**; Figure 4, Table 2 in Ref III). Additionally, the greater affinity of Indian Jews to Neolithic Iranian farmers than most Indian sub-continental populations (Figure 6, Ref III) links their migration from West Eurasia. Simultaneously, the higher ancestry of Neolithic Iranians in Mumbai Jews than in Kerala Jews is again suggestive of a more recent movement of Mumbai Jews to India than Kerala Jews.

Collectively, these results replicate the earlier discussed genetic variability between Mumbai Jews and Kerala Jews in the context of modern genomes and considerable mixing of Indian Jews with local South Asians successive to their migration from the Near East. Parsis' genetic analysis outcomes suggest a shared genetic pool for both the Indian and Pakistani Parsis, with the closest affinity to modern and ancient Middle Eastern populations and an evident one-way admixture from South Asia characterized by the gene flow from the South Asian females.

5.3. Reconstructing the genetic history of Tharu people: a linguistically diverse tribal group of Tarai region (Ref IV)

Tharu group is among the most prominent and linguistically diverse group that probably is indigenous to the Tarai region of southern Nepal and northern India. Tharu people are characterized by their higher affinity within themselves and to other local populations, which coexist with a substantial amount of East/Southeast Asian ancestry shown in earlier cultural and genetic studies (Crooke, 1896; Terrenato et al., 1988; Modiano et al., 1991; Passarino et al., 1992; Guneratne, 2002; Chhetri, 2005; Fornarino et al., 2009). To find the explanation for the tremendous genetic diversity shown by Tharus and if they have any common genetic heritage, we analyzed formerly published autosomal SNP data from Tharu groups of northern India (Consortium et al., 2009; Reich et al., 2009) and new high resolution uniparental genetic marker data of Tharus from Tarai region of Nepal and northern India.

5.3.1. Genome-wide variation of the Tharu groups

We examined the joint data of autosomal SNPs of Tharus from Uttar Pradesh and Uttarakhand, published earlier (Reich et al., 2009; Metspalu et al., 2011), relative to the broader data set across Eurasia (Supplementary Table 1 in Ref IV). We observed that although both groups of Tharus display the same level of genome-wide variability (Supplementary Table 1 in Ref IV), the genetic outline of Uttar Pradesh Tharu is distinct from their Uttarakhand counterpart. This was established by a greater intimacy with their corresponding nearby populations than they mutually are (Figure 2a/b, Supplementary Table 2 in Ref IV); and a pretty high F_{ST} differentiation between both groups that is two times higher than the

average mean pairwise distance between Pakistanis and Indian Dravidians. Analogous to these results, we observed a considerable difference regarding the East Asian ancestry component of both Tharu groups in ADMIXTURE analysis (Figure 2c in Ref IV). Such a superseding East Asian component among Uttarakhand Tharu individuals, unlike Uttar Pradesh Tharus, was also observed in the f_4 ancestry ratio (Table 1 in Ref IV), making them outliers like Kusunda, Spiti, Khasi, and Indian Austroasiatic people.

Consequently, we infer a recent admixture from East/Southeast Asia, rather than a deep sharing before the ancestral South Asians split from ancestral East Asians, which otherwise would have resulted in a uniformly distributed East Asian component among all contemporary populations of the region. In agreement with allele frequency-based analyses, Haplotype-based results also show a lesser chunk sharing between Tharu groups than their nearby ethnic groups and a similar mean chunk length sharing pattern for Uttar Pradesh Tharus as their nearby Indo-Europeans and Dravidians do. On the contrary, Tharus from Uttarakhand receives longer East/Southeast Asian chunks (**Figure 12a**). Thus, supporting a hypothesis that instead of a common genetic heritage, Tharu is an ethnolinguistic entity formed by assimilation of several tribes with different origins of substantial East/Southeast Asian ancestry.

5.3.2. Maternal and paternal lineage diversity in the Tharu

Uniparental results of the study found Tharu groups, in general, close to northern Indian groups with a slight shift towards populations from Southeast Asia and Tibet in PCA plots (**Figure 12b/c**). It suggests that the East Asian-related ancestry, present among people of Nepal and Tharu, most likely arrived from the north over the Himalayas, thus agreeing with the perception of earlier studies (Gayden et al., 2007; Fornarino et al., 2009; Wang et al., 2012). Interestingly, we observed a substantial variation in maternal and paternal lineages among Tharu groups, harmonious with the autosomal results. However, in contrast to the autosomal data analyses, we established a nominal amount of shared ancestry among Tharus, reflected by a more common East and Southeast Asian-related lineage sharing between Tharu groups (**Figure 12d/e**). The common East and Southeast Asian-related lineages display a Tharu group-specific cline from east to west, almost negligible among other resident groups of local Uttar Pradesh state (Table 2, Supplementary Table 5/6, in Ref IV). Nevertheless, we observed a higher diversity in South Asian-specific lineages among Tharus and other Nepali people, compared to that of East Asian-specific lineages, in the context of both uniparental loci (Supplementary Table 7 in Ref IV), supporting likely extensive interbreeding of Tharu people with local residents (Wang et al., 2012).

We unveiled a deeply rooted coalescent time of the rare M43 lineage, akin to other South Asian M hg subclans, coupled with the presence of highly variant M43 lineage among populations of Nepal and North India that is absent among South Indians and people of East/Southeast Asia (Supplementary Figure 5, Table 5

in Ref IV). Accordingly, we propose the origin of M43 lineage in the North Indian region, representing a current common maternal descent between all Tharu groups. The observed sharing of paternal East Asia derived O3a2c1-M134 hg between both Tharu groups could be a result of dispersal from Tibet and Southeast Asia to northern Indian states via Tarai of Nepal (Gayden et al., 2007; Thangaraj et al., 2008; Wang et al., 2012). It is mainly due to the widespread presence of the ancestral lineage of this hg among East and Southeast Asians (He et al., 2012). Y-STR results (Supplementary Figure 6, Table 4 in Ref IV) claim the involvement of several lineages in the migration, rather than a simple founder effect, that possibly happened due to a wide-ranging movement of people in the Holocene period.

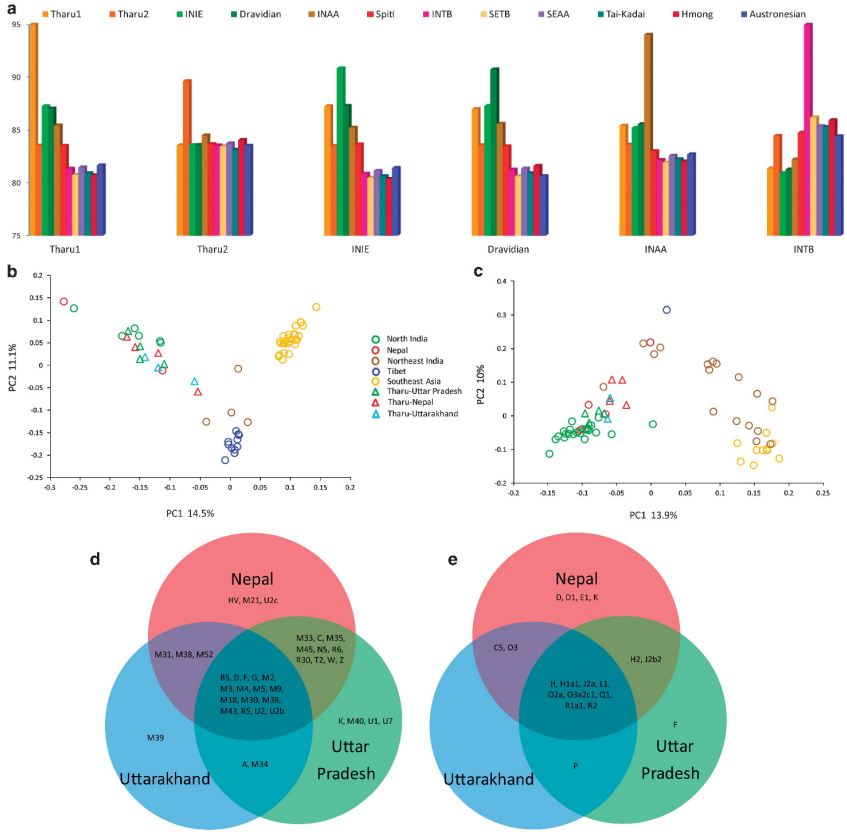

Figure 12. Haplotype sharing and haploid genetic diversity among Tharu and neighboring groups.

(a) The mean sharing of various chunk lengths among different ethnolinguistic South and East/Southeast Asian groups. Principal component analysis (PCA) based on mtDNA **(b)** and Y-chromosome **(c)** relative frequencies of haplogroups within populations. The mtDNA **(d)** and Y-chromosomal **(e)** haplogroup sharing among the different Tharu groups. INAA, Indian Austroasiatic; INIE, Indian Indo-European; INTB, Indian Tibeto-Burman; SEAA, Southeast Asian Austroasiatic; SETB, Southeast Asian Tibeto-Burman. Reprinted from Figure 3 of Ref IV, with permission from "Springer Nature".

Our study also detected declining frequency of the East/Southeast Asian related c.1540T>C allele of Ectodysplasin A receptor (EDAR) gene among Tharu groups, from Nepal to Uttar Pradesh (Supplementary Table 11, Ref IV). EDAR is a cell-surface receptor for ectodysplasin, a protein involved significantly in the development of ectodermal tissues (e.g., skin) and encoded by the EDAR gene humans (Aswegan et al., 1997; Monreal et al., 1999). Such a declining pattern combined with significant East Asian maternal and paternal lineages indicates a later spread of Tharus from the Tarai region after the episode of invasion from East Asia (Tables 2, 3; Supplementary Tables 8, 9, 11, Ref IV).

In summary, the higher genome-wide variability among Tharus from Uttar Pradesh and Uttarakhand supports the unification of Tharu as a cultural phenomenon rather than the demic. The East to the west decline of East Asian ancestry component of Tharu groups, together with a minor level of unique sharing in uniparental lineages, suggest a later spread of Tharus from the Tarai region, after the invasion from East Asia, which probably was followed by extensive mixing and integration with local populations.

6. CONCLUSIONS

- Substantial intra-regional genetic heterogeneity among contemporary populations of Indus Valley suggests a long-term population structure within the region that plausibly have persisted over thousands of years from the time of ancient Indus Civilization.
- We characterized population-specific demographic processes for culturally distinct groups of Northwest India such as Ror and Jat.
- The higher affinity of Ror to modern and ancient people living west of India, including in the ancient Swat Valley, most probably reflects the effect of their genetic continuum since the Bronze Age movement of northern Steppe herders.
- We identified a common genetic stock for both the Indian and Pakistan Parsis that has a close genetic affinity to modern and ancient West Eurasians, including Iranians.
- We revealed a male-biased migration of Parsis and initial admixture with local South Asian females before they turned into partial isolation and inbreeding.
- We detected a genetic variability between Jew groups of India in the context of both ancient and modern genomes and a pattern of considerable admixture of Indian Jews with local South Asian populations.
- We showed that the autosomal genetic analysis of different Tharu groups agrees largely with the model of predominantly cultural rather than demic unity of the Tharu.
- We revealed a substantial amount of East Asian genetic ancestry among Tharu peoples and a pattern of shared uniparental lineages that suggest the spread of the Tharu from Tarai region after the influx of East Asian genetic component.
- We detected a likely northern Indian subcontinental origin for some genetic ancestry shared amongst the Tharu – revealed by exclusive sharing of the South Asia specific mtDNA hg M43.

SUMMARY IN ESTONIAN

Induse jõe oru inimeste, parside, India juutide ja Tharu hõimu geneetilise põlvnemise piiritlemine

Lõuna-Aasia, esmajoones India rahvaste geneetiline varieeruvus, sh selle kujunemine, on olnud ja on Tartu geneetikute töömaaks alates mõõdunud sajandi viimasest kümnendist. Koostöös kolleegidega Indiast ja teistes riikidest, on ilmunud ligikaudu 40 teaduslikku artiklit, sealjuures kaitstud neli doktoriväitekirja, toetudes esmajoones just Tartus läbiviidud uuringute tulemustele. On eraldi märkimisväärne, et neist dissertatsioonidest kahe autoriteks on olnud eestlased ja kahe puhul on olnud tegu Tartu Ülikoolis doktorantuuri läbinud noore India teadlasega. Käesolev dissertatsioon samal teemal on seega viiendaks. Millest niisugune tähelepanu? Esmajoones põhjusel, et Lõuna-Aasia oli tõenäoliselt üks tänapäeva inimese esimesi elupaiku pärast meie liigi ulatuslikku väljarännet Aafrikast 50–70 tuhande aasta eest, kus tänapäeval elab enam kui viiendik inimkonnast. Seega pole meie liigi geneetilise varieeruvuse kujunemise mõistmine, ammugi siis Euraasias, võimalik ilma püüdeta mõista Lõuna-Aasia rahvaste geneetilist ajalugu. Nii ongi eelnevad Tartu Ülikoolis läbiviidud uuringud haaranud Lõuna-Aasia ja lähipiirkondade geneetilise varieeruvuse kujunemist alates kümnete tuhandete aastate sügavusest kuni viimase mõnetuhande aasta taguste sündmusteni.

Käesolevasse väitekirja on koondatud valik dissertandi kaasautorlusel ilmunud artiklitest, mille väga üldise iseloomustusena võiks nimetada püüet paremini mõista India rahvaste geneetilist ajalugu alates asustuse suhteliselt hilisest, kuid seda enam põnevast perioodist – nn. Induse tsivilisatsioonist.

Pronksiaegne Induse oru tsivilisatsioon (IOT). IOT (3300–1300 eKr) oli üks antiikmaailma kolmest suurimast tsivilisatsioonist, Mesopotaamia ja Egiptuse kaasaegne, ja selle juured olid Lõuna-Aasia loodeosas. IOT mõistatusliku hääbumise põhjustasid kõige tõenäolisemalt kuivad klimaatilised tingimused, mille tõttu IOT inimesed liikusid itta India Gangese tasandiku poole. Pakistan paikneb Induse jõe läänekaldal, samas kui Induse idakallas, tuntud kui varase veedade India (~2000–600 eKr) idapoolse Induse oru osa, hõlmab ajaloolise Kurukshetra ja kõrval paikneva Loode-India, kus elab mitmeid pikka aega püsinud etnilisi rühmi. India loode- ja põhjaosa on olnud kui inimeste liikumiste galerii eelajaloolisest ajast hiljutise minevikuni, mis tõi Indiasse Euraasia teistest osadest sisserändajaid, kes segunesid kohalikega ja jätsid endast jäljed tänapäeva India inimestesse.

Lisaks teevad kasti/hõimupõhine sotsiaalne struktuur ja grupisisese abiellumise tava Lõuna-Aasia praegused elanikud populatsioonigeneetiliste uuringute jaoks ainulaadseks. See erandlikkus on suur eelis, mis on piirkonna demograafilise ajaloo ja toimunud populatsioonidevaheliste kokkupuudete mõju selgitamisel tohutult abiks. Kuna Lõuna-Aasia on osutunud inimeste segamisnõuks, on parem arusaamine alaesindatud regionaalsete populatsioonide geneetilisest arhitektuurist vajalik selleks, et mõista tänast rahvastikku kujundanud demograafilist dünaamikat.

Tehnilised edusammud DNA eraldamisel iidsetest proovidest ning uued statistilised tööriistad tänapäeva ja vanade genoomsete andmete analüüsimiseks teevad populatsioonigeneetikast olulise vahendi inimese populatsiooni ajaloo seninägematul tasemel lahti mõtestamisel. Selliseid statistilisi tööriistu saab kasutada kindlate populatsioonide kogu genoomi kiibiandmete analüüsimiseks, koos nende mitokondriaalse DNA (mtDNA) ja Y-kromosoomi (chrY) andmetega, et teha kindlaks just selle populatsiooni ajalugu. Selliste populatsioonigeneetika tööriistade rakendamine paljude lõuna-aasialaste genoomsete andmete analüüsiks varem avaldatud globaalsete genoomsete andmete kontekstis võib lahti harutada segaseid arusaamu tänapäeva Induse oru, Põhja-India ja ka ajaloolisel ajal Lõuna-Aasiasse sisse rännanute mineviku populatsioonidünaamikast.

Et saada paremat ettekujutust Lõuna-Aasia ajaloolisest populatsioonidünaamikast, uurisime mõnda tänapäeva populatsiooni varasemates uuringutes alaesindatud piirkondadest, peamiselt Loode- ja Põhja-Indiast.

Lühidalt suurendab doktoritöö oluliselt varasemaid teadmisi Lõuna-Aasia rahvastiku dünaamikast, toetudes erinevate haploidsete ja diploidsete genoomsete markerite analüüsidele, kuid peamiselt ülegenoomsetele SNP-dele.

Suuremalt jaolt uurib doktoritöö tänapäeva rahvaste geneetilist varieeruvust Loode-India ajaloolises Induse oru piirkonnas, mõnda ajaloolisel ajal Lähis-Idast sisse rännanud gruppi, ja üht Põhja-India Tarai piirkonna hõimu, mis on geneetiline isolaat. Esiteks on doktoritöö eesmärk täiendada tänapäeva Induse oru inimeste alaesindatud geneetilist struktuuri, saades ja analüüsides uusi andmeid praegustelt populatsioonidelt, mille kohalolu Loode-Indias on pikalt dokumenteeritud (Ref I). Uuring käsitles tänapäeva populatsioonide ülegenoomset varieeruvust maailma tänapäeva populatsioonide ja ka kättesaadavate iidsete indiviidide geneetilise varieeruvuse kontekstis. Sellest selgus tänapäeva Induse oru populatsioonide peen struktureeritus ja geneetiline heterogeensus. Seejärel keskendub doktoritöö Lõuna-Aasia parsi ja juudi rahvusrühmadele, kes rändasid ajaloolisel ajal Lähis-Idast Indiasse, peamiselt, kuid mitte ainult põhjaossa. Uuringute eesmärk oli mõista parside ja juutide geneetilist struktuuri tänapäeva ja iidsete indiviidide kontekstis ning seda, kuidas nad pärast rännet segunesid kohalike inimestega (Ref II/III). Ref II analüüsis uniparentaalseid ja autosomaalseid andmeid, leides Lõuna-Aasia parsidel ühise geneetilise tausta, mis on lähedane iraanlastega, ning isolatsiooni ja sugulusristumise mustri. Samas analüüsis Ref III parside ja India juutide geneetilist varieeruvust võrreldes kättesaadavate iidsete genoomidega Lääne-Euraasiast, leides tänapäeva genoomidest ilmnevate mustrite kordumise. Viimaks käsitleb doktoritöö Tharu rahvusrühma, vana etnilist isolaati Põhja-India Tarai piirkonnast, uurimaks selle päritolu ning rühmasisest ja rühmadevahelist geneetilist varieeruvust võrreldes nende naabritega (Ref IV). Uuring avastas ülegenoomsetes analüüsides Tharu rühmade vahelise geneetilise erinevuse ning Tharu rühmade teataval määral ühise põlvnemise uniparentaalsetes andmetes, viidates nii sellele, et Tharud on kujunenud kultuurilise, mitte geneetilise ühtesulamise tagajärjel.

Kokkuvõttes rõhutab käesolev doktoritöö järgmisi punkte, täiendades oluliselt arusaamist Lõuna-Aasia populatsioonidünaamikast, keskendudes peamiselt, kuid mitte ainult India põhja- ja loodeosale.

- Märkimisväärne regioonisisene geneetiline heterogeensus Induse oru kaasaegsete populatsioonide seas viitab pikaaegsele populatsiooni struktureeritusele piirkonnas, mis võib olla püsinud tuhandeid aastaid, iidse Induse tsivilisatsiooni ajast saadik.
- Iseloomustasime Loode-India kultuuriliselt eristuvate rühmade nagu Ror ja Jat populatsioonispetsiifilisi demograafilisi protsesse.
- Rori suurem sarnasus tänapäeva ja iidsete inimestega, kes elavad/elasid Indiast lääne pool, sh iidses Swati orus, peegeldab kõige tõenäolisemalt nende geneetilist jätkuvust põhjapoolse stepi karjakasvatajate pronksiaegsest liikumisest saadik.
- Tegime kindlaks India ja Pakistani parside ühise geneetilise liini ning nende lähedase geneetilise sarnasuse tänapäeva ja iidsete lääne-euraaslastega, sh tänapäeva ja neoliitiliste iraanlastega, mis on seotud nende juurtega Pärsias – tänapäeva Iraanis.
- Näitasime parside rände kallutatust meeste poole, millele järgnes algne ühesuunaline segenemine kohalike Lõuna-Aasia naistega enne nende osalisse isolatsiooni ja sugulusristumise juurde jäämist
- Tuvastasime India juutide seas geneetilise varieeruvuse võrreldes iidsete genoomidega, mis oli märgatav ka tänapäeva genoomide kontekstis ning millele järgnes India juutide märkimisväärne segunemine neid ümbritsevate kohalike Lõuna-Aasia populatsioonidega.
- Näitasime, et erinevate Tharu rühmade autosomaalne geneetiline analüüs toetab üldiselt Tharude peamiselt kultuurilise, mitte geneetilise ühtsuse mudelit.
- Avastasime ka jälgi Tharu rahvaste jagatud ühisest põlvnemisest, eriti nende uniparentaalsetes lookustes, ning näitasime, et nende Ida-Aasia geneetiline komponent on vanem kui Tharude levik Indias.
- Leidsime kõigil Tharu rühmadel iidse haruldase mtDNA liini M43 ning paigutasime M43 tõenäolise päritolupiirkonna Põhja-Indiasse, kuna M43 leidub suure varieeruvusega nepaallaste ja põhja-indialaste hulgas, kuid see puudub Ida-Aasia, Kagu-Aasia ja teiste Lõuna-Aasia populatsioonide seas.

Harappa (Induse oru) ja naaberpopulatsioonide ulatuslikud uuringud on hädavajalikud Euraasia inimevolutsiooni ja etnolingvistilise ajaloo täielikuks rekonstruktsiooniks. Meie töö on pealegi osutanud, et mõnel Harappa piirkonna populatsioonil on ootamatult hiljutisem ühine põlvnemine Euroopa (Eesti) populatsioonidega, mis tõenäoliselt pärineb ühisest allikast stepivööndis. Need tähelepanekud võivad olla olulised uurijaile, nägemaks geene, mis määravad erinevaid Euroopa (Eesti) ja India populatsioonidele ühiseid fenotüübilisi tunnuseid. Lähis-Idast pärinevate ajaloolise aja migrantide, juutide ja parside detailne geneetiline analüüs parandab oluliselt arusaamist populatsiooni taseme interaktsioonidest, mis kujundasid nii migrantide (parsid, India juudid) kui ka lõuna-aasialaste

genoome. Parside populatsiooni isolatsiooni ja sugulusristumise muster võib leida rakendust haruldaste haiguste meditsiinigeneetilistes uuringutes. Tharu etnilise isolaadi teke on olnud segadusttekitav küsimus, mis käesolevas doktoritöös lahendatakse, näidates erinevate Tharu rühmade vahel märkimisväärset geneetilist erinevust.

Kokkuvõttes toovad need tulemused mõnevõrra suuremat selgust erinevate Lõuna-Aasia rahvaste evolutsioonilisse ajalukku, sealjuures konteksts, mis seob Lõuna-Aasiat Euraasia kui tevikuga ja samas illustreerib mõnede valitud representatiivsete näidete kujul maailma rahvaste demograafilisele ajaloole iseloomulikke üldiseid protsesse, mis on seotud rännetega.

REFERENCES

Abbi, A. 2006. Endangered languages of the Andaman Islands (Lincom Europa) Adler, D.S., Wilkinson, K.N., Blockley, S., Mark, D.F., Pinhasi, R., et al. 2014. Early Levallois technology and the Lower to Middle Paleolithic transition in the Southern Caucasus. *Science* 345: 1609–1613.Agarwal, V. 2003. A Reply to Michael Witzel's 'Ein Fremdling im Rgveda.' *Journal of Indo-European Studies* 31: 107–185. Akhilesh, K., Pappu, S., Rajapara, H.M., Gunnell, Y., Shukla, A.D., and Singhvi, A.K. 2018. Early Middle Palaeolithic culture in India around 385–172 ka reframes Out of Africa models. *Nature* 554: 97–101.Albrechtsen, A., Korneliussen, T.S., Moltke, I., Hansen, T. van O., Nielsen, F.C., and Nielsen, R. 2009. Relatedness mapping and tracts of relatedness for genome-wide data in the presence of linkage disequilibrium. *Genetic Epidemiology* 33: 266–274.Alexander, D.H., Novembre, J., and Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19.

Allchin, B., and Allchin, F.R. 1997. Origins of a civilization: the prehistory and early archaeology of South Asia (Viking Adult).

Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., and Vinner, L. 2015. Population genomics of bronze age Eurasia. *Nature* 522: 167–172.

Álvarez-Alonso, D. 2014. First Neanderthal settlements in northern Iberia: The Acheulean and the emergence of Mousterian technology in the Cantabrian region. *Quaternary International* 326: 288–306.

Ambrose, S.H. 1998. Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans. *Journal of Human Evolution* 34: 623–651.

Ammerman, A.J., and Biagi, P. 2003. The widening harvest: the Neolithic transition in Europe: looking back, looking forward (Archaeological Inst of Amer).

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., and Sanger, F. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290: 457–465.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23: 147–147.

Armitage, S.J., Jasim, S.A., Marks, A.E., Parker, A.G., Usik, V.I., and Uerpmann, H.-P. 2011. The Southern Route "Out of Africa": Evidence for an Early Expansion of Modern Humans into Arabia. *Science* 331: 453–456.

Arunkumar, G., Wei, L., Kavitha, V.J., Syama, A., Arun, V.S., Sathua, S., Sahoo, R., Balakrishnan, R., Riba, T., and Chakravarthy, J. 2015. A late Neolithic expansion of Y chromosomal haplogroup O2a1-M95 from east to west. *Journal of Systematics and Evolution* 53: 546–560.

ASI 2003. Rock Shelters of Bhimbetka. 233.

Aswegan, A.L., Josephson, K.D., Mowbray, R., Pauli, R.M., Spritz, R.A., and Williams, M.S. 1997. Autosomal dominant hypohidrotic ectodermal dysplasia in a large family. *Am. J. Med. Genet.* 72: 462–467.

Athreya, S. 2007. Was Homo heidelbergensis in South Asia? A test using the Narmada fossil from central India. In The Evolution and History of Human Populations in South Asia, (Springer), pp. 137–170.

Atzmon, G., Hao, L., Pe'er, I., Velez, C., Pearlman, A., Palamara, P.F., Morrow, B., Friedman, E., Oddoux, C., and Burns, E. 2010. Abraham's children in the genome era:

major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. *The American Journal of Human Genetics* 86: 850–859.

Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., and McVean, G.A. 2015. A global reference for human genetic variation. *Nature* 526.

Auton, A., Bryc, K., Boyko, A.R., Lohmueller, K.E., Novembre, J., et al. 2009. Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res.* 19: 795–803.

Ayub, Q., Mezzavilla, M., Pagani, L., Haber, M., Mohyuddin, A., Khaliq, S., Mehdi, S.Q., and Tyler-Smith, C. 2015a. The Kalash genetic isolate: ancient divergence, drift, and selection. *Am J Hum Genet* 96.

Ayub, Q., Mezzavilla, M., Pagani, L., Haber, M., Mohyuddin, A., Khaliq, S., Mehdi, S.Q., and Tyler-Smith, C. 2015b. The Kalash Genetic Isolate: Ancient Divergence, Drift, and Selection. *Am J Hum Genet* 96: 775–783.

Bae, C.J., Douka, K., and Petraglia, M.D. 2017. On the origin of modern humans: Asian perspectives. *Science* 358: eaai9067.

Bae, C.J., Wang, W., Zhao, J., Huang, S., Tian, F., and Shen, G. 2014. Modern human teeth from late Pleistocene Luna cave (Guangxi, China). *Quaternary International* 354: 169–183.

Bamshad, M., Kivisild, T., Watkins, W.S., Dixon, M.E., Ricker, C.E., Rao, B.B., Naidu, J.M., Prasad, B.R., Reddy, P.G., and Rasanayagam, A. 2001. Genetic evidence on the origins of Indian caste populations. *Genome Research* 11: 994–1004.

Bamshad, M.J., Watkins, W.S., Dixon, M.E., Jorde, L.B., Rao, B.B., Naidu, J.M., Prasad, B.R., Rasanayagam, A., and Hammer, M.F. 1998. Female gene flow stratifies Hindu castes. *Nature* 395: 651.

Baran, Y., Pasaniuc, B., Sankararaman, S., Torgerson, D.G., Gignoux, C., et al. 2012. Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics* 28: 1359–1367.

Basu, A., Mukherjee, N., Roy, S., Sengupta, S., Banerjee, S., Chakraborty, M., Dey, B., Roy, M., Roy, B., and Bhattacharyya, N.P. 2003. Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res* 13.

Basu, A., Sarkar-Roy, N., and Majumder, P.P. 2016. Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. *Proc Natl Acad Sci U S A* 113.

Beck, R.B., and Beck, R.B. 2005. World history: Patterns of interaction (McDougal Littell Evanston, IL).

Behar, D.M., Metspalu, E., Kivisild, T., Rosset, S., Tzur, S., Hadid, Y., Yudkovsky, G., Rosengarten, D., Pereira, L., and Amorim, A. 2008a. Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. *Plos One* 3: e2062.

Behar, D.M., Metspalu, M., Baran, Y., Kopelman, N.M., Yunusbayev, B., et al. 2013. No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. *Hum. Biol.* 85: 859–900.

Behar, D.M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N.M., Kivisild, T., Torroni, A., and Villems, R. 2012. A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from its Root. *The American Journal of Human Genetics* 90: 675–684.

Behar, D.M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S., and Comas, D. 2008b. The dawn of human matrilineal diversity. *The American Journal of Human Genetics* 82: 1130–1140.

Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., and Yudkovsky, G. 2010. The genome-wide structure of the Jewish people. *Nature* 466: 238.

Benazzi, S., Douka, K., Fornai, C., Bauer, C.C., Kullmer, O., et al. 2011. Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* 479: 525–528.

Benedetti, G. 2014. The chronology of Puranic kings and Rigvedic rishis in comparison with the phases of the Sindhu-Sarasvati civilization. *Sindu-Sarasvati Civilization: New Perspectives* 220–246.

Bhan, S. 1972. Prehistoric archaeology of the Sarasvati and the drisadvati valleys Haryana C 2300 1500 BC.

Bittles, A.H. 2002. Endogamy, consanguinity and community genetics. *Journal of Genetics* 81: 91–98.

Blazek, V. 1999. Elam: a bridge between Ancient Near East and Dravidian India. *Archaeology and Language IV. Language Change and Cultural Transformation. London, New York: Routledge* 48–78.

Blench, R. 2014. Tracking the origins of African slaves in the Indian Ocean through personal names: the evidence of Sumatra records. *Afriques on East Africa and the Indian Ocean* 2–18.

Boivin, N. 2007. Anthropological, historical, archaeological and genetic perspectives on the origins of caste in South Asia. In The Evolution and History of Human Populations in South Asia, (Springer), pp. 341–361.

Bradburd, G.S., Ralph, P.L., and Coop, G.M. 2016. A Spatial Framework for Understanding Population Structure and Admixture. *PLOS Genetics* 12: e1005703.

Bräuer, G. 2008. The origin of modern anatomy: by speciation or intraspecific evolution? *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews* 17: 22–37.

Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., et al. 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325: 318–321.

Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., Reynolds, A., Ostrer, H., Mezey, J.G., and Bustamante, C.D. 2012. PCAdmix: Principal Components-Based Assignment of Ancestry Along Each Chromosome in Individuals with Admixed Ancestry from Two or More Populations. *Hbio* 84: 343–364.

Brooke, J.L. 2014. Climate change and the course of global history: a rough journey (Cambridge University Press).

Broushaki, F., Thomas, M.G., Link, V., López, S., van Dorp, L., Kirsanow, K., Hofmanová, Z., Diekmann, Y., Cassidy, L.M., and Díez-del-Molino, D. 2016. Early Neolithic genomes from the eastern Fertile Crescent. *Science* 353: 499–503.

Brown, W.M., George, M., and Wilson, A.C. 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* 76: 1967–1971.

Browning, B.L., and Browning, S.R. 2011. A Fast, Powerful Method for Detecting Identity by Descent. *The American Journal of Human Genetics* 88: 173–182.

Browning, B.L., and Browning, S.R. 2013. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194: 459–471.

Browning, S.R., and Browning, B.L. 2015. Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. *The American Journal of Human Genetics* 97: 404–418.

Browning, S.R., and Browning, B.L. 2010. High-Resolution Detection of Identity by Descent in Unrelated Individuals. *The American Journal of Human Genetics* 86: 526–539.

Browning, S.R., and Browning, B.L. 2007. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *American Journal of Human Genetics* 81: 1084.

Browning, S.R., Browning, B.L., Zhou, Y., Tucci, S., and Akey, J.M. 2018. Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell* 173: 53-61.e9.

Brucato, N., Fernandes, V., Kusuma, P., Černý, V., Mulligan, C.J., et al. 2019. Evidence of Austronesian Genetic Lineages in East Africa and South Arabia: Complex Dispersal from Madagascar and Southeast Asia. *Genome Biol Evol* 11: 748–758.

Brucato, N., Kusuma, P., Beaujard, P., Sudoyo, H., Cox, M.P., and Ricaut, F.-X. 2017. Genomic admixture tracks pulses of economic activity over 2,000 years in the Indian Ocean trading network. *Scientific Reports* 7: 2919.

Bryc, K., Auton, A., Nelson, M.R., Oksenberg, J.R., Hauser, S.L., et al. 2010. Genome-wide patterns of population structure and admixture in West Africans and African Americans. *PNAS* 107: 786–791.

Busby, G.B.J., Hellenthal, G., Montinaro, F., Tofanelli, S., Bulayeva, K., et al. 2015. The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. *Current Biology* 25: 2518–2526.

Calafell, F., and Larmuseau, M.H. 2017. The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. *Human Genetics* 136: 559–573.

Callaway, E. 2020. Ancient human genomes shed new light on East Asia's history. *Nature*.

Cameron, D., Patnaik, R., and Sahni, A. 2004. The phylogenetic significance of the Middle Pleistocene Narmada hominin cranium from central India. *International Journal of Osteoarchaeology* 14: 419–447.

Cann, R.L., Stoneking, M., and Wilson, A.C. 1987. Mitochondrial DNA and human evolution. *Nature* 325: 31–36.

Carelli, V. 2015. Keeping in shape the dogma of mitochondrial DNA maternal inheritance. *PLoS Genetics* 11.

Carpenter, M.L., Buenrostro, J.D., Valdiosera, C., Schroeder, H., Allentoft, M.E., et al. 2013. Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *Am J Hum Genet* 93: 852–864.

Cavalli-Sforza, L.L. 1996. The spread of agriculture and nomadic pastoralism: insights from the genetics, linguistics and archaeology. *The Origins and Spread of Agriculture and Pastoralism in Eurasia* 51–69.

Cavalli-Sforza, L.L., Cavalli-Sforza, L., Menozzi, P., and Piazza, A. 1994. The history and geography of human genes (Princeton university press).

Chahal, S.M., and Papiha, S.S. 1981. A population genetic study of the Jat Sikhs, Punjab, India. *J. Indian Anthrop. Soc* 16: 251–260.

Chakrabarty, D.K. 2009. India: an archaeological history: palaeolithic beginnings to early historic foundations (Oxford University Press).

Chakraborty, R., and Weiss, K.M. 1988. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proceedings of the National Academy of Sciences* 85: 9119–9123.

Chan, E.K.F., Timmermann, A., Baldi, B.F., Moore, A.E., Lyons, R.J., et al. 2019. Human origins in a southern African palaeo-wetland and first migrations. *Nature* 575: 185–189.

Chandrasekar, A., Kumar, S., Sreenath, J., Sarkar, B.N., Urade, B.P., et al. 2009. Updating Phylogeny of Mitochondrial DNA Macrohaplogroup M in India: Dispersal of Modern Human in South Asian Corridor. *PLoS ONE* 4: 1–13.

Chatterjee, A., Ray, J.S., Shukla, A.D., and Pande, K. 2019. On the existence of a perennial river in the Harappan heartland. *Sci Rep* 9: 1–7.

Chaubey, G., Kadian, A., Bala, S., and Rao, V.R. 2015. Genetic affinity of the Bhil, Kol and Gond mentioned in epic ramayana. *PloS One* 10: e0127655.

Chaubey, G., Karmin, M., Metspalu, E., Metspalu, M., Selvi-Rani, D., Singh, V.K., Parik, J., Solnik, A., Naidu, B.P., and Kumar, A. 2008a. Phylogeography of mtDNA haplogroup R7 in the Indian peninsula. *BMC Evolutionary Biology* 8: 227.

Chaubey, G., Metspalu, M., Choi, Y., Magi, R., Romero, I.G., Soares, P., van Oven, M., Behar, D.M., Rootsi, S., and Hudjashov, G. 2011a. Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Mol Biol Evol* 28.

Chaubey, G., Metspalu, M., Choi, Y., Mägi, R., Romero, I.G., Soares, P., van Oven, M., Behar, D.M., Rootsi, S., and Hudjashov, G. 2011b. Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Molecular Biology and Evolution* 28: 1013–1024.

Chaubey, G., Metspalu, M., Karmin, M., Thangaraj, K., Rootsi, S., Parik, J., Solnik, A., Rani, D.S., Singh, V.K., and Naidu, B.P. 2008b. Language shift by indigenous population: a model genetic study in South Asia. *International Journal of Human Genetics* 8: 41–50.

Chaubey, G., Metspalu, M., Kivisild, T., and Villems, R. 2007. Peopling of South Asia: investigating the caste-tribe continuum in India. *BioEssays* 29.

Chaubey, G., Singh, M., Rai, N., Kariappa, M., Singh, K., et al. 2016. Genetic affinities of the Jewish populations of India. *Scientific Reports* 6: 19166.

Chaubey, G., Tamang, R., Pennarun, E., Dubey, P., Rai, N., Upadhyay, R.K., Meena, R.P., Patel, J.R., van Driem, G., and Thangaraj, K. 2017. Reconstructing the population history of the largest tribe of India: the Dravidian speaking Gond. *European Journal of Human Genetics* 25: 493–498.

Chen, L., Wolf, A.B., Fu, W., Li, L., and Akey, J.M. 2020. Identifying and Interpreting Apparent Neanderthal Ancestry in African Individuals. *Cell* 180: 677-687.e16.

Chhetri, R.B. 2005. The plight of the Tharu Kamaiyas in Nepal: A review of the social, economic and political facets. *Occasional Papers in Sociology and Anthropology* 9: 22–46.

Chiang, C.W.K., Marcus, J.H., Sidore, C., Biddanda, A., Al-Asadi, H., et al. 2018. Genomic history of the Sardinian population. *Nat Genet* 50: 1426–1434.

Chidbhavananda, S. 1965. The Bhagavad Gita: Original Stanzas (Tapovanam Publishing House).

Churchhouse, C., and Marchini, J. 2013. Multiway Admixture Deconvolution Using Phased or Unphased Ancestral Panels. *Genetic Epidemiology* 37: 1–12.

Clarkson, C., Harris, C., Li, B., Neudorf, C.M., Roberts, R.G., et al. 2020. Human occupation of northern India spans the Toba super-eruption ~74,000 years ago. *Nature Communications* 11: 1–10.

Clarkson, C., Jacobs, Z., Marwick, B., Fullagar, R., Wallis, L., Smith, M., Roberts, R.G., Hayes, E., Lowe, K., and Carah, X. 2017. Human occupation of northern Australia by 65,000 years ago. *Nature* 547: 306.

Clarkson, C., Jones, S., and Harris, C. 2012. Continuity and change in the lithic industries of the Jurreru Valley, India, before and after the Toba eruption. *Quaternary International* 258: 165–179.

Clarkson, C., Petraglia, M., Korisettar, R., Haslam, M., Boivin, N., Crowther, A., Ditchfield, P., Fuller, D., Miracle, P., and Harris, C. 2009. The oldest and longest enduring microlithic sequence in India: 35 000 years of modern human occupation and change at the Jwalapuram Locality 9 rockshelter. *Antiquity* 83: 326–348.

Clift, P.D., Carter, A., Giosan, L., Durcan, J., Duller, G.A.T., et al. 2012. U-Pb zircon dating evidence for a Pleistocene Sarasvati River and capture of the Yamuna River. *Geology* 40: 211–214.

Cohen, T., Levene, C., Yodfat, Y., Fidel, J., Friedlander, Y., Steinberg, A., Brautbar, C., and Opitz, J.M. 1980. Genetic studies on Cochin Jews in Israel: 1. Population data, blood groups, isoenzymes, and HLA determinants. *American Journal of Medical Genetics* 6: 61–73.

Consortium, H.P.-.A.S., Abdulla, M.A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S.K., Calacal, G.C., Chaurasia, A., Chen, C.H., and Chen, J. 2009. Mapping human genetic diversity in Asia. *Science* 326:.

Consortium, T.Y.C. 2002. A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome Res.* 12: 339–348.

Cordaux, R., Aunger, R., Bentley, G., Nasidze, I., Sirajuddin, S.M., and Stoneking, M. 2004a. Independent origins of Indian caste and tribal paternal lineages. *Current Biology* 14: 231–235.

Cordaux, R., Deepa, E., Vishwanathan, H., and Stoneking, M. 2004b. Genetic evidence for the demic diffusion of agriculture to India. *Science* 304: 1125–1125.

Cristofaro, J.D., Pennarun, E., Mazières, S., Myres, N.M., Lin, A.A., et al. 2013. Afghan Hindu Kush: Where Eurasian Sub-Continent Gene Flows Converge. *PLOS ONE* 8: e76748.

Crooke, W. 1896. The tribes and castes of the North-Western Provinces and Oudh (Office of the superintendent of government printing).

Cummings, V., Jordan, P., and Zvelebil, M. 2014. The Oxford handbook of the archaeology and anthropology of hunter-gatherers (OUP Oxford).

Damgaard, P. de B., Martiniano, R., Kamm, J., Moreno-Mayar, J.V., Kroonen, G., et al. 2018. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* eaar7711.

Dange, S.S. 1984. The Bhagavata Purana: Mytho-Social Study (Ajanta Publications).

Dannemann, M., and Racimo, F. 2018. Something old, something borrowed: admixture and adaptation in human evolution. *Current Opinion in Genetics & Development* 53: 1–8.

Debortoli, G., Abbatangelo, C., Ceballos, F., Fortes-Lima, C., Norton, H.L., Ozarkar, S., Parra, E.J., and Jonnalagadda, M. 2020. Novel insights on demographic history of tribal and caste groups from West Maharashtra (India) using genome-wide data. *Scientific Reports* 10: 10075.

Delson, E. 2019. An early dispersal of modern humans from Africa to Greece. *Nature*.

deMenocal, P.B., and Stringer, C. 2016. Human migration: Climate and the peopling of the world. *Nature* 538: 49–50.

Demeter, F., Shackelford, L., Westaway, K., Barnes, L., Duringer, P., Ponche, J.-L., Dumoncel, J., Sénégas, F., Sayavongkhamdy, T., and Zhao, J.-X. 2017. Early modern humans from Tam Pà Ling, Laos: Fossil review and perspectives. *Current Anthropology* 58: S527–S538.

Deraniyagala, S.U. 1992. The Prehistory of Sri Lanka (Department of Archaeological Survey).
Derenko, M., Malyarchuk, B., Bahmanimehr, A., Denisova, G., Perkova, M., Farjadian, S., and Yepiskoposyan, L. 2013. Complete Mitochondrial DNA Diversity in Iranians. *PLOS ONE* 8: e80673.
Détroit, F., Mijares, A.S., Corny, J., Daver, G., Zanolli, C., Dizon, E., Robles, E., Grün, R., and Piper, P.J. 2019. A new species of Homo from the Late Pleistocene of the Philippines. *Nature* 568: 181–186.
Diamond, J., and Bellwood, P. 2003. Farmers and their languages: the first expansions. *Science* 300:.
Diffloth, G. 2009. More on Dvaravati Old Mon. p.
Dixit, Y., Hodell, D.A., and Petrie, C.A. 2014. Abrupt weakening of the summer monsoon in northwest India~ 4100 yr ago. *Geology* 42: 339–342.
Dumanski, J.P., Lambert, J.-C., Rasi, C., Giedraitis, V., Davies, H., Grenier-Boley, B., Lindgren, C.M., Campion, D., Dufouil, C., and Initiative, T.E.A.D. 2016. Mosaic loss of chromosome Y in blood is associated with Alzheimer disease. *The American Journal of Human Genetics* 98: 1208–1219.
Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. 2011. Testing for Ancient Admixture between Closely Related Populations. *Mol Biol Evol* 28: 2239–2252.
Durvasula, A., and Sankararaman, S. 2019. Recovering signals of ghost archaic introgression in African populations. *BioRxiv* 285734.
Eaaswarkhanth, M., Dubey, B., Meganathan, P.R., Ravesh, Z., Khan, F.A., Singh, L., Thangaraj, K., and Haque, I. 2009. Diverse genetic origin of Indian Muslims: evidence from autosomal STR loci. *J Hum Genet* 54: 340–348.
Eaaswarkhanth, M., Haque, I., Ravesh, Z., Romero, I.G., Meganathan, P.R., et al. 2010. Traces of sub-Saharan and Middle Eastern lineages in Indian Muslim populations. *Eur J Hum Genet* 18: 354–363.
Eaaswarkhanth, M., Pathak, A.K., Ongaro, L., Montinaro, F., Hebbar, P., Alsmadi, O., Metspalu, M., Al-Mulla, F., and Thanaraj, T.A. 2021. Unraveling a fine-scale high genetic heterogeneity and recent continental connections of an Arabian Peninsula population. *European Journal of Human Genetics* 1–13.
Eberhard, David, M., Simons, G.F., and Fennig, C.D. 2019. Etnhnologue: Languages of the World.
Ehrlich, M.A. 2009. Encyclopedia of the Jewish diaspora: origins, experiences, and culture (Abc-clio).
Eller, E. 2001. Effects of ascertainment bias on recovering human demographic history. *Hum. Biol.* 73: 411–427.
Eskandari-Nasab, E., Moghadampour, M., Najibi, H., and Hadadi-Fishani, M. 2014. Investigation of CTLA-4 and CD86 gene polymorphisms in Iranian patients with brucellosis infection. *Microbiology and Immunology* 58: 135–141.
Fedorova, S.A., Reidla, M., Metspalu, E., Metspalu, M., Rootsi, S., et al. 2013. Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol. Biol.* 13: 127.
Fernandes, E. 2011. The Last Jews of Kerala (Granta Books).
Fernandes, V., Brucato, N., Ferreira, J.C., Pedro, N., Cavadas, B., Ricaut, F.-X., Alshamali, F., and Pereira, L. 2019. Genome-Wide Characterization of Arabian Peninsula Populations: Shedding Light on the History of a Fundamental Bridge between Continents. *Molecular Biology and Evolution* 36: 575–586.

Field, J.S., Petraglia, M.D., and Lahr, M.M. 2007. The southern dispersal hypothesis and the South Asian archaeological record: examination of dispersal routes through GIS analysis. *Journal of Anthropological Archaeology* 26: 88–108.

Filosto, M., Mancuso, M., Vives-Bauza, C., Vilà, M.R., Shanske, S., Hirano, M., Andreu, A.L., and DiMauro, S. 2003. Lack of paternal inheritance of muscle mitochondrial DNA in sporadic mitochondrial myopathies. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 54: 524–526.

Foley, R., and Lahr, M.M. 1997. Mode 3 technologies and the evolution of modern humans. *Cambridge Archaeological Journal* 7: 3–36.

Fornarino, S., Pala, M., Battaglia, V., Maranta, R., Achilli, A., Modiano, G., Torroni, A., Semino, O., and Santachiara-Benerecetti, S.A. 2009. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. *BMC Evol Biol* 9: 154.

Forsberg, L.A., Rasi, C., Malmqvist, N., Davies, H., Pasupulati, S., Pakalapati, G., Sandgren, J., de Ståhl, T.D., Zaghlool, A., and Giedraitis, V. 2014. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nature Genetics* 46: 624–628.

François, O., Currat, M., Ray, N., Han, E., Excoffier, L., and Novembre, J. 2010. Principal component analysis under population genetic models of range expansion and admixture. *Mol. Biol. Evol.* 27: 1257–1268.

Fregel, R., Méndez, F.L., Bokbot, Y., Martín-Socas, D., Camalich-Massieu, M.D., et al. 2018. Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *PNAS* 115: 6774–6779.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., et al. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514: 445–449.

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Pääbo, S. 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U.S.A.* 110: 2223–2227.

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., et al. 2016. The genetic history of Ice Age Europe. *Nature* 534: 200.

Fuller, D.Q. 2003. An agricultural perspective on Dravidian historical linguistics: archaeological crop packages, livestock and Dravidian crop vocabulary. (McDonald Institute for Archaeological Research), p.

Fuller, D.Q. 2007. Non-human genetics, agricultural origins and historical linguistics in South Asia in The Evolution and History of Human Populations in South Asia (eds Petraglia, M. D. & Allchin, B.) 393–443 (Springer: Netherlands, 2007).

Gallego Romero, I., Basu Mallick, C., Liebert, A., Crivellaro, F., Chaubey, G., et al. 2012. Herders of Indian and European Cattle Share Their Predominant Allele for Lactase Persistence. *Molecular Biology and Evolution* 29: 249–260.

Gallego-Llorente, M., Connell, S., Jones, E.R., Merrett, D.C., Jeon, Y., et al. 2016. The genetics of an early Neolithic pastoralist from the Zagros, Iran. *Scientific Reports* 6: 31326.

Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., and Anders, A. 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications* 5: 1–9.

Gangal, K., Sarson, G.R., and Shukurov, A. 2014. The near-eastern roots of the Neolithic in South Asia. *PLoS One* 9.

Gangal, K., Vahia, M.N., and Adhikari, R. 2010. Spatio-temporal analysis of the Indus urbanization. *Current Science (00113891)* 98: 846–852.

Gansauge, M.-T., and Meyer, M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc* 8: 737–748.

Gaur, A.S., and Vora, K. 1999. Ancient shorelines of Gujarat, India, during the Indus civilization (Late Mid-Holocene): A study based on archaeological evidences. *Current Science* 180–185.

Gayden, T., Cadenas, A.M., Regueiro, M., Singh, N.B., Zhivotovsky, L.A., Underhill, P.A., Cavalli-Sforza, L.L., and Herrera, R.J. 2007. The Himalayas as a directional barrier to gene flow. *The American Journal of Human Genetics* 80: 884–894.

Ghose, B., Kar, A., and Husain, Z. 1979. Lost courses of the Saraswati river in the great Indian desert – new evidence from Landsat imagery. *Geog. J.* 145.

Giosan, L., Clift, P.D., Macklin, M.G., Fuller, D.Q., Constantinescu, S., et al. 2012a. Fluvial landscapes of the Harappan civilization. *PNAS* 109: E1688–E1694.

Giosan, L., Clift, P.D., Macklin, M.G., Fuller, D.Q., Constantinescu, S., et al. 2012b. Fluvial landscapes of the Harappan civilization. *Proceedings of the National Academy of Sciences* 109: E1688–E1694.

Giosan, L., Fuller, D.Q., Nicoll, K., Flad, R.K., and Clift, P.D. 2013. Climates, Landscapes, and Civilizations. (Washington: Wiley).

Giosan, L., Orsi, W.D., Coolen, M., Wuchter, C., Dunlea, A.G., et al. 2018. Neoglacial climate anomalies and the Harappan metamorphosis. *Clim. Past* 14: 1669–1686.

Gravel, S. 2012. Population Genetics Models of Local Ancestry. *Genetics* 191: 607–619.

Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F., Gibbs, R.A., Bustamante, C.D., and 1000 Genomes Project 2011. Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences* 108: 11983–11988.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., et al. 2010. A Draft Sequence of the Neandertal Genome. *Science* 328: 710–722.

Griffith, R.T.H. 2017. The Rig Veda: Complete (Illustrated) (CreateSpace Independent Publishing Platform).

Gronau, I., Hubisz, M.J., Gulko, B., Danko, C.G., and Siepel, A. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nat. Genet.* 43: 1031–1034.

Groucutt, H.S., Grün, R., Zalmout, I.A.S., Drake, N.A., Armitage, S.J., et al. 2018. Homo sapiens in Arabia by 85,000 years ago. *Nature Ecology & Evolution* 2: 800.

Groucutt, H.S., Petraglia, M.D., Bailey, G., Scerri, E.M., Parton, A., Clark-Balzan, L., Jennings, R.P., Lewis, L., Blinkhorn, J., and Drake, N.A. 2015. Rethinking the dispersal of Homo sapiens out of Africa. *Evolutionary Anthropology: Issues, News, and Reviews* 24: 149–164.

Grugni, V., Battaglia, V., Kashani, B.H., Parolo, S., Al-Zahery, N., et al. 2012. Ancient Migratory Events in the Middle East: New Clues from the Y-Chromosome Variation of Modern Iranians. *PLOS ONE* 7: e41252.

Grün, R., Stringer, C., McDermott, F., Nathan, R., Porat, N., Robertson, S., Taylor, L., Mortimer, G., Eggins, S., and McCulloch, M. 2005. U-series and ESR analyses of bones and teeth relating to the human burials from Skhul. *Journal of Human Evolution* 49: 316–334.

GTEx Consortium 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348: 648–660.

Guneratne, A. 2002. Many tongues, one people: The making of Tharu identity in Nepal (Cornell University Press).

Gupta, A.K., Anderson, D.M., Pandey, D.N., and Singhvi, A.K. 2006. Adaptation and human migration, and evidence of agriculture coincident with changes in the Indian summer monsoon during the Holocene. *CURRENT SCIENCE-BANGALORE*-90: 1082.

Gusev, A., Lowe, J.K., Stoffel, M., Daly, M.J., Altshuler, D., Breslow, J.L., Friedman, J.M., and Pe'er, I. 2009. Whole population, genome-wide mapping of hidden relatedness. *Genome Research* 19: 318–326.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., and Stewardson, K. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522: 207–211.

Haber, M., Gauguier, D., Youhanna, S., Patterson, N., Moorjani, P., Botigué, L.R., Platt, D.E., Matisoo-Smith, E., Soria-Hernanz, D.F., and Wells, R.S. 2013. Genome-wide diversity in the levant reveals recent structuring by culture. *PLoS Genetics* 9: e1003316.

Hagström, E., Freyer, C., Battersby, B.J., Stewart, J.B., and Larsson, N.-G. 2014. No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline. *Nucleic Acids Res.* 42: 1111–1116.

Hajdinjak, M., Mafessoni, F., Skov, L., Vernot, B., Hübner, A., et al. 2021. Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry. *Nature* 592: 253–257.

Harney, É., Nayak, A., Patterson, N., Joglekar, P., Mushrif-Tripathy, V., et al. 2019. Ancient DNA from the skeletons of Roopkund Lake reveals Mediterranean migrants in India. *Nature Communications* 10: 3670.

Harris, A.M., and DeGiorgio, M. 2017. Admixture and Ancestry Inference from Ancient and Modern Samples through Measures of Population Genetic Drift. *Hum. Biol.* 89: 21–46.

Harris, K., and Nielsen, R. 2013. Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genetics* 9: e1003521.

Harvati, K., Röding, C., Bosman, A.M., Karakostis, F.A., Grün, R., et al. 2019. Apidima Cave fossils provide earliest evidence of Homo sapiens in Eurasia. *Nature* 1.

Haslam, M., Clarkson, C., Petraglia, M., Korisettar, R., Janardhana, B., Boivin, N., Ditchfield, P., Jones, S., and Mackay, A. 2010. Indian lithic technology prior to the 74,000 BP Toba super-eruption: searching for an early modern human signature.

He, J.-D., Peng, M.-S., Quang, H.H., Dang, K.P., Trieu, A.V., Wu, S.-F., Jin, J.-Q., Murphy, R.W., Yao, Y.-G., and Zhang, Y.-P. 2012. Patrilineal perspective on the Austronesian diffusion in Mainland Southeast Asia. *PLoS One* 7: e36437.

Helgason, A., Einarsson, A.W., Guðmundsdóttir, V.B., Sigurðsson, Á., Gunnarsdóttir, E.D., Jagadeesan, A., Ebenesersdóttir, S.S., Kong, A., and Stefánsson, K. 2015. The Y-chromosome point mutation rate in humans. *Nature Genetics* 47: 453–457.

Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., and Myers, S. 2014. A Genetic Atlas of Human Admixture History. *Science* 343: 747–751.

Hemphill, B.E. 1991. Biological adaptations and affinities of Bronze Age Harappans. *Harappa Excavations 1986–1990: A Multidisciplinary Approach to Third Millennium Urbanism (Monographs in World Archaeology No. 3)* 137–182.

Hershkovitz, I., Duval, M., Grün, R., Mercier, N., Valladas, H., et al. 2018a. Response to Comment on "The earliest modern humans outside Africa." *Science* 362.

Hershkovitz, I., Weber, G.W., Quam, R., Duval, M., Grün, R., Kinsley, L., Ayalon, A., Bar-Matthews, M., Valladas, H., and Mercier, N. 2018b. The earliest modern humans outside Africa. *Science* 359: 456–459.

Heyer, E., Chaix, R., Pavard, S., and Austerlitz, F. 2012. Sex-specific demographic behaviours that shape human genomic variation. *Molecular Ecology* 21: 597–612.
Higham, T., Compton, T., Stringer, C., Jacobi, R., Shapiro, B., et al. 2011. The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* 479: 521–524.
Hinnells, J.R. 2007. Parsis in India and the Diaspora in the Twentieth Century and Beyond. In Parsis in India and the Diaspora, (Routledge), pp. 269–290.
Hinnells, J.R. 1994. The modern Zoroastrian diaspora. In Migration: The Asian Experience, (Springer), pp. 56–82.
Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-del-Molino, D., van Dorp, L., López, S., Kousathanas, A., and Link, V. 2016. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences* 113: 6886–6891.
Hoggart, C.J., Shriver, M.D., Kittles, R.A., Clayton, D.G., and McKeigue, P.M. 2004. Design and Analysis of Admixture Mapping Studies. *The American Journal of Human Genetics* 74: 965–978.
Holsinger, K.E., and Weir, B.S. 2009. Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat Rev Genet* 10: 639–650.
Hovers, E., and Kuhn, S. 2007. Transitions before the transition: evolution and stability in the Middle Paleolithic and Middle Stone Age (Springer Science & Business Media).
Hsieh, P., Woerner, A.E., Wall, J.D., Lachance, J., Tishkoff, S.A., Gutenkunst, R.N., and Hammer, M.F. 2016. Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African Pygmies. *Genome Res.* 26: 291–300.
Hublin, J.-J., Ben-Ncer, A., Bailey, S.E., Freidline, S.E., Neubauer, S., Skinner, M.M., Bergmann, I., Le Cabec, A., Benazzi, S., and Harvati, K. 2017. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. *Nature* 546: 289.
Hutchison, C.A., Newbold, J.E., Potter, S.S., and Edgell, M.H. 1974. Maternal inheritance of mammalian mitochondrial DNA. *Nature* 251: 536–538.
INDIA, P.O.M.P.I. 2011. Census of India 2011 Provisional Population Totals.
Indian Genome Variation Consortium 2008. Genetic landscape of the people of India: a canvas for disease gene exploration. *J Genet* 87: 3–20.
International Hap Map 2005. A haplotype map of the human genome. *Nature* 437: 1299–1320.
International HapMap Consortium 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* 467: 52–58.
Irwin, D.E. 2002. Phylogeographic breaks without geographic barriers to gene flow. *Evolution* 56: 2383–2394.
Israel, R.R. 2002. The Jews of India: their story/Rachel Rukmini Israel (Mosaic Books).
Jacobs, G.S., Hudjashov, G., Saag, L., Kusuma, P., Darusallam, C.C., et al. 2019. Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell* 177: 1010-1021.e32.
James, H.A., Petraglia, M., deBeaune, S., Dennell, R., Kivisild, T., Korisettar, R., Lukacs, J., Misra, V., James, H.A., and Petraglia, M. 2005. Modern human origins and the evolution of behavior in the later Pleistocene record of South Asia. *Current Anthropology* 46: S3–S27.
Jarrige, C. 1995. Mehrgarh: Field reports 1974–1985, from Neolithic times to the Indus Civilization (Dept. of Culture and Tourism, Govt. of Sindh; in collaboration with the French Ministry of Foreign Affairs).

Jarrige, J.-F. 1986. Excavations at Mehrgarh-Nausharo. *Pakistan Archaeology* 10: 63–131.
Jarrige, J.F. 2006. Mehrgarh Neolithic, Pragdhara No. 18.
Järve, M., Saag, L., Scheib, C.L., Pathak, A.K., Montinaro, F., Pagani, L., Flores, R., Guellil, M., Saag, L., and Tambets, K. 2019. Shifts in the Genetic Landscape of the Western Eurasian Steppe Associated with the Beginning and End of the Scythian Dominance. *Current Biology* 29: 2430–2441.
Jin, W., Xu, S., Wang, H., Yu, Y., Shen, Y., Wu, B., and Jin, L. 2012. Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Res.* 22: 519–527.
Jobling, M., Hollox, E., Hurles, M., Kivisild, T., and Tyler-Smith, C. 2014. Human Evolutionary Genetics Second Edition. *Human Evolutionary Genetics, Second Edition* 601–608.
Jobling, M.A., Pandya, A., and Tyler-Smith, C. 1997. The Y chromosome in forensic analysis and paternity testing. *International Journal of Legal Medicine* 110: 118–124.
Jobling, M.A., and Tyler-Smith, C. 1995. Fathers and sons: the Y chromosome and human evolution. *Trends in Genetics* 11: 449–456.
Jobling, M.A., and Tyler-Smith, C. 2017. Human Y-chromosome variation in the genome-sequencing era. *Nat Rev Genet* 18: 485–497.
Jobling, M.A., and Tyler-Smith, C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetics* 4: 598–612.
Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., et al. 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature Communications* 6: 8912.
Jones, S.C. 2007. The Toba supervolcanic eruption: Tephra-fall deposits in India and paleoanthropological implications. In The Evolution and History of Human Populations in South Asia, (Springer), pp. 173–200.
Jonnalagadda, M., Ozarkar, S., and Mushrif-Tripathy, V. 2011. Population affinities of Parsis in the Indian subcontinent. *International Journal of Osteoarchaeology* 21: 103–110.
Joseph, T. 2017. Who built the Indus Valley civilisation? *The Hindu*.
Kaifu, Y., Izuho, M., Goebel, T., Sato, H., and Ono, A. 2014. Emergence and diversity of modern human behavior in Paleolithic Asia (Texas A&M University Press).
Kar, A., and Ghose, B. 1984. The Drishadvati river system of India: an assessment and new findings. *Geographical Journal* 221–229.
Karkal, M. 1975. Marriage among Parsis. *Demography India* 4: 128–145.
Karkal, M. 1984. Survey of Parsi Population of Greater Bombay-1982 (Trustees of the Parsi Punchayet Funds and Properties).
Karmin, M., Saag, L., Vicente, M., Sayres, M.A.W., Järve, M., et al. 2015. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 25: 459–466.
Kashimada, K., and Koopman, P. 2010. Sry: the master switch in mammalian sex determination. *Development* 137: 3921–3930.
Katz, N. 2000. Who are the Jews of India? (Univ of California Press).
Kennedy, K.A. 2001. Middle and late Pleistocene hominids of south Asia. In Humanity from African Naissance to Coming Millennia, (Firenze University Press; Witwatersrand university press), pp. 1000–1008.
Kennedy, K.A., and Deraniyagala, S.U. 1989. Fossil remains of 28,000-year-old hominids from Sri Lanka. *Current Anthropology* 30: 394–399.

Kenoyer, J., and Meadow, R. 2016. Excavations at Harappa, 1986–2010: New Insights on the Indus Civilization and Harappan Burial Traditions. *A Companion to South Asia in the Past* 145–168.

Kenoyer, J.M. 2006. Cultures and societies of the Indus tradition. *Historical Roots in the Making of 'the Aryan', National Book Trust, New Delhi* 21–49.

Kenoyer, J.M. 1998. Kenoyer, J. M. Ancient Cities of the Indus Valley Civilization. (Oxford University, Oxford, 1998).

Kenoyer, J.M. 2008. Kenoyer, J. M. in The Ancient City: New Perspectives in the Old and New World (eds J. Marcus & J. A. Sablof) 85–109 (SAR, Santa Fe, NM, 2008).

Kenoyer, M., and Meadow, R.H. 2010. Inscribed objects from Harappa excavations 1986–2007. *Corpus of Indus Seals and Inscriptions 3: New Material Untraced Objects and Collections Outside India and Pakistan Part 1: Mohenjo-Daro and Harappa*.

Khan, I., and Sinha, R. 2019. Discovering 'buried' channels of the Palaeo-Yamuna river in NW India using geophysical evidence: Implications for major drainage reorganization and linkage to the Harappan Civilization. *Journal of Applied Geophysics* 167: 128–139.

Kidd, J.M., Gravel, S., Byrnes, J., Moreno-Estrada, A., Musharoff, S., et al. 2012. Population Genetic Inference from Personal Genome Data: Impact of Ancestry and Admixture on Human Genomic Variation. *The American Journal of Human Genetics* 91: 660–671.

Kirsanow, K., and Burger, J. 2012. Ancient human DNA. *Annals of Anatomy-Anatomischer Anzeiger* 194: 121–132.

Kivisild, T. 2015. Maternal ancestry and population history from whole mitochondrial genomes. *Investigative Genetics* 6: 3.

Kivisild, T., Bamshad, M.J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., Laos, S., Parik, J., Watkins, W.S., and Dixon, M.E. 1999. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9.

Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H.-V., and Stepanov, V. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *The American Journal of Human Genetics* 72: 313–332.

de Knijff, P. 2000. Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *The American Journal of Human Genetics* 67: 1055–1061.

Koirala, S., Bhattacharya, S., Paudel, I.S., Yadav, B.N., Shah, S., and Baral, P. 2012. A Comparative Anthropometric Study of Mongoloid and Tharu Ethnic Races in Eastern Nepal.

Korisettar, R. 2017. Beyond stones and more stones: defining Indian prehistoric archaeology (MYTHIC SOCIETY).

Krause, J., Orlando, L., Serre, D., Viola, B., Prüfer, K., Richards, M.P., Hublin, J.-J., Hänni, C., Derevianko, A.P., and Pääbo, S. 2007. Neanderthals in central Asia and Siberia. *Nature* 449: 902–904.

Krishnamoorthy, S., and Audinarayana, N. 2001. Trends in consanguinity in South India. *Journal of Biosocial Science* 33: 185–197.

Kuhlwilm, M., Gronau, I., Hubisz, M.J., de Filippo, C., Prado-Martinez, J., et al. 2016. Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* 530: 429–433.

Kumar, M. 1978. Archaeology of Ambala and Kurukshetra Districts (Haryana). *Unpublished Ph. D. Thesis. Kurukshetra University, Kurukshetra* 148–149.
Kumar, V., Reddy, A.N., Babu, J.P., Rao, T.N., Langstieh, B.T., Thangaraj, K., Reddy, A.G., Singh, L., and Reddy, B.M. 2007. Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evol Biol* 7.
Kurland, C.G., Collins, L.J., and Penny, D. 2006. Genomics and the irreducible nature of eukaryote cells. *Science* 312: 1011–1014.
Kurukshetra District 2009. History of India: Ancient Kurukshetra.
Kushniarevich, A., Utevska, O., Chuhryaeva, M., Agdzhoyan, A., Dibirova, K., et al. 2015. Genetic Heritage of the Balto-Slavic Speaking Populations: A Synthesis of Autosomal, Mitochondrial and Y-Chromosomal Data. *PLOS ONE* 10: e0135820.
Lachance, J., and Tishkoff, S.A. 2013. SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *BioEssays* 35: 780–786.
Lahr, M.M., and Foley, R. 1994. Multiple dispersals and modern human origins. *Evolutionary Anthropology: Issues, News, and Reviews* 3: 48–60.
Lahr, M.M.N., and Foley, R.A. 1998. Towards a Theory of Modern Human Origins: Geography, Demography, and Diversity in Recent Human Evolution. *Yearbook of Physical Anthropology* 41: 137–176.
Lal, P. 1980. Mahabharata of Vyasa (Asia Book Corporation of America).
Landsteiner, K. 1900. Zur Kenntniss der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. Zbl. Bakt. 27: 357–362. 1901 Ueber Agglutinationserscheinungen normalen menschlichen Blutes. *Wien. Klin. Wschr* 14: 1132–1134.
Landsteiner, K., and Wiener, A.S. 1940. An agglutinable factor in human blood recognized by immune sera for rhesus blood. *Proceedings of the Society for Experimental Biology and Medicine* 43: 223–223.
Laso-Jadart, R., Harmant, C., Quach, H., Zidane, N., Tyler-Smith, C., Mehdi, Q., Ayub, Q., Quintana-Murci, L., and Patin, E. 2017. The Genetic Legacy of the Indian Ocean Slave Trade: Recent Admixture and Post-admixture Selection in the Makranis of Pakistan. *The American Journal of Human Genetics* 101: 977–984.
Lawler, A. 2008. Boring no more, a trade-savvy Indus emerges. *Science* 320: 1276–1281.
Lawler, A. 2013. Mohenjo-Daro's New Story. *Archaeology* 66: 32–37.
Lawson, D.J., Dorp, L. van, and Falush, D. 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat Commun* 9: 1–11.
Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. 2012. Inference of population structure using dense haplotype data. *PLoS Genetics* 8: e1002453.
Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., and Sirak, K. 2016. Genomic insights into the origin of farming in the ancient Near East. *Nature* 536: 419–424.
Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513: 409.
Lee, W.J., Shinde, V., Kim, Y.J., Woo, E.J., Jadhav, N., et al. 2019. Craniofacial reconstruction of the Indus Valley Civilization individuals found at 4500-year-old Rakhigarhi cemetery. *Anat Sci Int*.
Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., et al. 2015. The fine-scale genetic structure of the British population. *Nature* 519: 309.
Levine, P., and Stetson, R.E. 1939. An unusual case of intra-group agglutination. *Journal of the American Medical Association* 113: 126–127.

Li, H., and Durbin, R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475: 493.

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., et al. 2008. Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. *Science* 319: 1100–1104.

Lipson, M., Loh, P.-R., Levin, A., Reich, D., Patterson, N., and Berger, B. 2013. Efficient Moment-Based Inference of Admixture Parameters and Sources of Gene Flow. *Mol Biol Evol* 30: 1788–1802.

Lipson, M., Ribot, I., Mallick, S., Rohland, N., Olalde, I., et al. 2020. Ancient West African foragers in the context of African population history. *Nature* 577: 665–670.

Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., et al. 2017. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* 551: 368.

Liu, W., Martinón-Torres, M., Cai, Y., Xing, S., Tong, H., Pei, S., Sier, M.J., Wu, X., Edwards, R.L., and Cheng, H. 2015. The earliest unequivocally modern humans in southern China. *Nature* 526: 696.

Llamas, B., Willerslev, E., and Orlando, L. 2017. Human evolution: a tale from ancient genomes. *Phil. Trans. R. Soc. B* 372: 20150484.

Llorente, M.G., Jones, E.R., Eriksson, A., Siska, V., Arthur, K., Arthur, J., Curtis, M., Stock, J.T., Coltorti, M., and Pieruccini, P. 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* 350: 820–822.

Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. 2013. Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. *Genetics* 193: 1233–1254.

López, S., Thomas, M.G., van Dorp, L., Ansari-Pour, N., Stewart, S., et al. 2017. The Genetic Legacy of Zoroastrianism in Iran and India: Insights into Population Structure, Gene Flow, and Selection. *The American Journal of Human Genetics* 101: 353–368.

Luo, S., Valencia, C.A., Zhang, J., Lee, N.-C., Slone, J., Gui, B., Wang, X., Li, Z., Dell, S., and Brown, J. 2018. Biparental inheritance of mitochondrial DNA in humans. *Proceedings of the National Academy of Sciences* 115: 13039–13044.

Luo, S., Valencia, C.A., Zhang, J., Lee, N.-C., Slone, J., Gui, B., Wang, X., Li, Z., Dell, S., and Brown, J. 2019. Reply to Lutz-Bonengel et al.: Biparental mtDNA transmission is unlikely to be the result of nuclear mitochondrial DNA segments. *Proceedings of the National Academy of Sciences* 116: 1823–1824.

Lutz-Bonengel, S., and Parson, W. 2019. No further evidence for paternal leakage of mitochondrial DNA in humans yet. *Proceedings of the National Academy of Sciences* 116: 1821–1822.

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., and Cruciani, F. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308: 1034–1036.

MacDonald, G. 2011. Potential influence of the Pacific Ocean on the Indian summer monsoon and Harappan decline. *Quaternary International* 229: 140–148.

Madella, M., and Fuller, D.Q. 2006. Palaeoecology and the Harappan Civilisation of South Asia: a reconsideration. *Quaternary Science Reviews* 25: 1283–1301.

Mafessoni, F., Grote, S., Filippo, C. de, Slon, V., Kolobova, K.A., et al. 2020. A high-coverage Neandertal genome from Chagyrskaya Cave. *PNAS*.

Majumder, P.P. 2010. The human genetic history of South Asia. *Current Biology* 20: R184–R187.
Majumder, P.P., and Basu, A. 2015. A genomic view of the peopling and population structure of India. *Cold Spring Harbor Perspectives in Biology* 7: a008540.
Malaspinas, A.-S., Westaway, M.C., Muller, C., Sousa, V.C., Lao, O., Alves, I., Bergström, A., Athanasiadis, G., Cheng, J.Y., and Crawford, J.E. 2016. A genomic history of Aboriginal Australia. *Nature* 538: 207.
Mallick, C.B., Iliescu, F.M., Möls, M., Hill, S., Tamang, R., Chaubey, G., Goto, R., Ho, S.Y., Romero, I.G., and Crivellaro, F. 2013. The light skin allele of SLC24A5 in South Asians and Europeans shares identity by descent. *PLoS Genet* 9: e1003912.
Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., and Tandon, A. 2016. The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature* 538: 201.
Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. 2013. RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *The American Journal of Human Genetics* 93: 278–288.
Mardis, E.R. 2011. A decade's perspective on DNA sequencing technology. *Nature* 470: 198–203.
Maricic, T., Whitten, M., and Pääbo, S. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* 5: e14004.
Mason, P.A., Matheson, E.C., Hall, A.G., and Lightowlers, R.N. 2003. Mismatch repair activity in mammalian mitochondria. *Nucleic Acids Research* 31: 1052–1058.
Mastana, S.S., and Papiha, S.S. 1992. Origin of the Romany gypsies—genetic evidence. *Zeitschrift Für Morphologie Und Anthropologie* 43–51.
Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., et al. 2018. The genomic history of southeastern Europe. *Nature* 555: 197–203.
Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., and Novak, M. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528: 499–503.
McAlpin, D.W. 1981. Proto-Elamo-Dravidian: The evidence and its implications. *Transactions of the American Philosophical Society* 71: 1–155.
McBrearty, S. 2003. Patterns of technological change at the origin of Homo sapiens. *Before Farming* 2003: 1–6.
McClish, M., and Olivelle, P. 2012. The Artha__stra: Selections from the Classic Indian Work on Statecraft (M.S. Koṭhārī).
McDougall, I., Brown, F.H., and Fleagle, J.G. 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* 433: 733.
McElreavey, K., and Quintana-Murci, L. 2005. A population genetics perspective of the Indus Valley through uniparentally-inherited markers. *Annals of Human Biology* 32: 154–162.
McIntosh, J. 2008. The ancient Indus Valley: new perspectives (Santa Barbara, Calif: ABC-CLIO).
McLean, J. 1999. Conservation and the Impact of Relocation on the Tharus of Chitwan, Nepal. 8.
McVean, G. 2009. A Genealogical Interpretation of Principal Components Analysis. *PLOS Genetics* 5: e1000686.
McVean, G.A., Altshuler (Co-Chair), D.M., Durbin (Co-Chair), R.M., Abecasis, G.R., Bentley, D.R., et al. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56–65.

Meadow, R.H. 1991. Harappa excavations 1986–1990: a multidisciplinary approach to Third Millenium urbanism (Prehistory Pr).
Mehdi, S.M., Pant, N.C., Saini, H.S., Mujtaba, S.A.I., and Pande, P. 2016. Identification of palaeochannel configuration in the Saraswati River basin in parts of Haryana and Rajasthan, India, through digital remote sensing and GIS. *Episodes* 39.
Mellars, P. 2006a. A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature* 439: 931.
Mellars, P. 2006b. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* 313: 796–800.
Mellars, P. 2006c. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proceedings of the National Academy of Sciences* 103: 9381–9386.
Mellars, P., Gori, K.C., Carr, M., Soares, P.A., and Richards, M.B. 2013. Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. *Proc Natl Acad Sci U S A* 110: 10699–10704.
Merriwether, D.A., Clark, A.G., Ballinger, S.W., Schurr, T.G., Soodyall, H., Jenkins, T., Sherry, S.T., and Wallace, D.C. 1991. The structure of human mitochondrial DNA variation. *J. Mol. Evol.* 33: 543–555.
Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., and Gilbert, M.T. 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5.
Metspalu, M., Mondal, M., and Chaubey, G. 2018. The genetic makings of South Asia. *Current Opinion in Genetics & Development* 53: 128–133.
Metspalu, M., Romero, I.G., Yunusbayev, B., Chaubey, G., Mallick, C.B., Hudjashov, G., Nelis, M., Mägi, R., Metspalu, E., and Remm, M. 2011. Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *The American Journal of Human Genetics* 89: 731–744.
Meyer, M., Arsuaga, J.-L., de Filippo, C., Nagel, S., Aximu-Petri, A., et al. 2016. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* 531: 504–507.
Meyer, M., and Kircher, M. 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb Protoc* 2010: pdb.prot5448.
Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., et al. 2012. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* 338: 222–226.
Michaels, G.S., Hauswirth, W.W., and Laipis, P.J. 1982. Mitochondrial DNA copy number in bovine oocytes and somatic cells. *Dev. Biol.* 94: 246–251.
Michel, V., Valladas, H., Shen, G., Wang, W., Zhao, J., Shen, C.-C., Valensi, P., and Bae, C.J. 2016. The earliest modern Homo sapiens in China. *Journal of Human Evolution* 101: 1e104.
Mikhed, Y., Daiber, A., and Steven, S. 2015. Mitochondrial Oxidative Stress, Mitochondrial DNA Damage and Their Role in Age-Related Vascular Dysfunction. *Int J Mol Sci* 16: 15918–15953.
Mineta, K., Goto, K., Gojobori, T., and Alkuraya, F.S. 2021. Population structure of indigenous inhabitants of Arabia. *PLOS Genetics* 17: e1009210.
Mirazón Lahr, M., and Foley, R.A. 1998. Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *American Journal*

of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists 107: 137–176.
Mirza, H.D.K. 1974. Outlines of Parsi history (Mirza).
Mishra, S., Chauhan, N., and Singhvi, A.K. 2013. Continuity of Microblade Technology in the Indian Subcontinent Since 45 ka: Implications for the Dispersal of Modern Humans. *PLoS One* 8.
Misra, S.S. 2004. The date of the Rigveda and the Aryan migration: fresh linguistic evidence. In The Indo-Aryan Controversy, (Routledge), pp. 193–245.
Misra, V.N. 1993. Indus civilization and the Rigvedic Saraswati. *South Asian Archaeology* 2: 511–525.
Misra, V.N. 2001a. Prehistoric human colonization of India. *J Biosci* 26.
Misra, V.N. 2001b. Prehistoric human colonization of India. *Journal of Biosciences* 26: 491–531.
Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V., and Tõrv, M. 2018. The genetic prehistory of the Baltic Sea region. *Nature Communications* 9: 1–11.
Modiano, G., Morpurgo, G., Terrenato, L., Novelletto, A., Di Rienzo, A., Colombo, B., Purpura, M., Mariani, M., Santachiara-Benerecetti, S., and Brega, A. 1991. Protection against malaria morbidity: near-fixation of the alpha-thalassemia gene in a Nepalese population. *Am. J. Hum. Genet.* 48: 390–397.
Mohsenpour, K., and Chandrasekar, A. 2019. Genetic affinity of Muslim population in South India based on HLA-DQB1 and relationship with other Indian Populations. *International Journal of Modern Anthropology* 2: 97–113.
Mohyuddin, A., and Mehdi, S. 2005. HLA analysis of the Parsi (Zoroastrian) population in Pakistan. *Tissue Antigens* 66: 691–695.
Molinaro, L., and Pagani, L. 2018. Human evolutionary history of Eastern Africa. *Current Opinion in Genetics & Development* 53: 134–139.
Mondal, M., Bergström, A., Xue, Y., Calafell, F., Laayouni, H., Casals, F., Majumder, P.P., Tyler-Smith, C., and Bertranpetit, J. 2017. Y-chromosomal sequences of diverse Indian populations and the ancestry of the Andamanese. *Hum Genet* 136: 499–510.
Mondal, M., Bertranpetit, J., and Lao, O. 2019. Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. *Nature Communications* 10: 246.
Mondal, M., Casals, F., Majumder, P.P., and Bertranpetit, J. 2018. Reply to 'No evidence for unknown archaic ancestry in South Asia.' *Nature Genetics* 50: 1637–1639.
Mondal, M., Casals, F., Xu, T., Dall'Olio, G.M., Pybus, M., Netea, M.G., Comas, D., Laayouni, H., Li, Q., and Majumder, P.P. 2016. Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nature Genetics* 48: 1066.
Monreal, A.W., Ferguson, B.M., Headon, D.J., Street, S.L., Overbeek, P.A., and Zonana, J. 1999. Mutations in the human homologue of mouse dl cause autosomal recessive and dominant hypohidrotic ectodermal dysplasia. *Nature Genetics* 22: 366–369.
Montinaro, F., Busby, G.B.J., Pascali, V.L., Myers, S., Hellenthal, G., and Capelli, C. 2015. Unravelling the hidden ancestry of American admixed populations. *Nature Communications* 6: 6596.
Moorjani, P., Patterson, N., Hirschhorn, J.N., Keinan, A., Hao, L., Atzmon, G., Burns, E., Ostrer, H., Price, A.L., and Reich, D. 2011. The History of African Gene Flow into Southern Europeans, Levantines, and Jews. *PLOS Genetics* 7: e1001373.

Moorjani, P., Thangaraj, K., Patterson, N., Lipson, M., Loh, P.R., Govindaraj, P., Berger, B., Reich, D., and Singh, L. 2013. Genetic evidence for recent population mixture in India. *Am J Hum Genet* 93.

Moreno-Mayar, J.V., Potter, B.A., Vinner, L., Steinrücken, M., Rasmussen, S., et al. 2018. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* 553: 203–207.

Moseley, C. 2008. Encyclopedia of the World's Endangered Languages (Routledge).

Mughal, M.R. 1997. Ancient Cholistan: archaeology and architecture (Ferozsons).

Muztar, B.K. 1978. Kurukshetra: Political and Cultural History.

Naegele, C.J. 2008. History and influence of law code of Manu.

Nakatsuka, N., Moorjani, P., Rai, N., Sarkar, B., Tandon, A., et al. 2017. The promise of discovering population-specific disease-associated genes in South Asia. *Nature Genetics* 49: 1403–1407.

Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., et al. 2019. The formation of human populations in South and Central Asia. *Science* 365.

Narasimhan, V.M., Patterson, N.J., Moorjani, P., Lazaridis, I., Mark, L., et al. 2018. The Genomic Formation of South and Central Asia. *BioRxiv* 292581.

Nath, A., Walimbe, S., Garge, T.M., Mushrif-Tripathy, V., Dehuri, R., and Malik, A. 2015. Harappan interments at Rakhigarhi, Haryana. *Man and Environment* 40.

Nelis, M., Esko, T., Mägi, R., Zimprich, F., Zimprich, A., et al. 2009. Genetic structure of Europeans: a view from the North-East. *PLoS ONE* 4: e5472.

Ng, A.Y., and Jordan, M.I. 2001. On discriminative vs. generative classifiers: a comparison of logistic regression and naïve bayes. *On Discriminative Vs. Generative Classifiers: A Comparison of Logistic Regression and Naive Bayes* 841–848.

Ngamphiw, C., Assawamakin, A., Xu, S., Shaw, P.J., Yang, J.O., Ghang, H., Bhak, J., Liu, E., Tongsima, S., and Consortium, and the H.P.-A.S. 2011. PanSNPdb: The Pan-Asian SNP Genotyping Database. *PLOS ONE* 6: e21451.

Nielsen, R., Akey, J.M., Jakobsson, M., Pritchard, J.K., Tishkoff, S., and Willerslev, E. 2017. Tracing the peopling of the world through genomics. *Nature* 541: 302–310.

Nordborg, M. 1998. On the probability of Neanderthal ancestry. *American Journal of Human Genetics* 63: 1237.

Novembre, J. 2016. Pritchard, Stephens, and Donnelly on Population Structure. *Genetics* 204: 391–393.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., et al. 2008. Genes mirror geography within Europe. *Nature* 456: 98–101.

Novembre, J., and Peter, B.M. 2016. Recent advances in the study of fine-scale population structure in humans. *Current Opinion in Genetics & Development* 41: 98–105.

Novembre, J., and Ramachandran, S. 2011. Perspectives on human population structure at the cusp of the sequencing era. *Annu Rev Genomics Hum Genet* 12: 245–274.

Novembre, J., and Stephens, M. 2008. Interpreting principal component analyses of spatial population genetic variation. *Nat Genet* 40.

O'Flaherty, W.D. 1981. The Rig Veda: an anthology: one hundred and eight hymns, selected, translated and annotated (Harmondsworth, Middlesex, England; New York, N.Y.: Penguin Books, 1981.).

Omberg, L., Salit, J., Hackett, N., Fuller, J., Matthew, R., Chouchane, L., Rodriguez-Flores, J.L., Bustamante, C., Crystal, R.G., and Mezey, J.G. 2012. Inferring genome-wide patterns of admixture in Qataris using fifty-five ancestral populations. *BMC Genetics* 13: 49.

Oppenheimer, S. 2012a. A single southern exit of modern humans from Africa: Before or after Toba? *Quaternary International* 258: 88–99.
Oppenheimer, S. 2012b. Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 770–784.
Oppenheimer, S. 2009. The great arc of dispersal of modern humans: Africa to Australia. *Quaternary International* 202: 2–13.
Orlando, L., Gilbert, M.T.P., and Willerslev, E. 2015. Reconstructing ancient genomes and epigenomes. *Nat. Rev. Genet.* 16: 395–408.
Pääbo, S. 2015. The diverse origins of the human gene pool. *Nature Reviews Genetics* 16: 313–314.
Padhukasahasram, B. 2014. Inferring ancestry from population genomic data and its applications. *Front. Genet.* 5.
Pagani, L., Colonna, V., Tyler-Smith, C., and Ayub, Q. 2017. An ethnolinguistic and genetic perspective on the origins of the Dravidian-speaking Brahui in Pakistan. *Man in India* 97: 267–277.
Pagani, L., Lawson, D.J., Jagoda, E., Mörseburg, A., Eriksson, A., Mitt, M., Clemente, F., Hudjashov, G., DeGiorgio, M., and Saag, L. 2016. Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* 538: 238–242.
Pagani, L., Schiffels, S., Gurdasani, D., Danecek, P., Scally, A., et al. 2015. Tracing the Route of Modern Humans out of Africa by Using 225 Human Genome Sequences from Ethiopians and Egyptians. *The American Journal of Human Genetics* 96: 986–991.
Page, D.C., Harper, M.E., Love, J., and Botstein, D. 1984. Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature* 311: 119–123.
Pal, Y., Sahai, B., Sood, R.K., and Agrawal, D.P. 1980. Remote sensing of the lost Sarasvati river. *Proc. Ind. Acad. Sci.* 89.
Palamara, P.F., Lencz, T., Darvasi, A., and Pe'er, I. 2012. Length Distributions of Identity by Descent Reveal Fine-Scale Demographic History. *The American Journal of Human Genetics* 91: 809–822.
Palamara, P.F., and Pe'er, I. 2013. Inference of historical migration rates via haplotype sharing. *Bioinformatics* 29: i180–i188.
Palanichamy, M. gounder, Sun, C., Agrawal, S., Bandelt, H.-J., Kong, Q.-P., Khan, F., Wang, C.-Y., Chaudhuri, T.K., Palla, V., and Zhang, Y.-P. 2004. Phylogeny of Mitochondrial DNA Macrohaplogroup N in India, Based on Complete Sequencing: Implications for the Peopling of South Asia. *The American Journal of Human Genetics* 75: 966–978.
Palsetia, J.S. 2001. The Parsis of India: preservation of identity in Bombay city (Brill).
Papiha, S., Mukherjee, B., Chahal, S., Malhotra, K., and Roberts, D. 1982. Genetic heterogeneity and population structure in north-west India. *Annals of Human Biology* 9: 235–251.
Pappu, S., Gunnell, Y., Akhilesh, K., Braucher, R., Taieb, M., Demory, F., and Thouveny, N. 2011. Early Pleistocene Presence of Acheulian Hominins in South India. *Science* 331: 1596–1599.
Parpola, A. 1988. The coming of the Aryans to Iran and India and the cultural and ethnic identity of the Dāsas. *Studia Orientalia Electronica* 64: 195–302.
Parpola, A. 1986. The Indus script: A challenging puzzle. *World Archaeology* 17: 399–419.

Parpola, A. 1979. The problem of the Indus script. *Essays in Indian Protohistory, BR, Delhi* 163–186.

Paşaniuc, B., Sankararaman, S., Kimmel, G., and Halperin, E. 2009. Inference of locus-specific ancestry in closely related populations. *Bioinformatics* 25: i213–i221.

Passarino, G., Semino, O., Modiano, G., and Santachiara-Benerecetti, A.S. 1993. COII/tRNA(Lys) intergenic 9-bp deletion and other mtDNA markers clearly reveal that the Tharus (southern Nepal) have Oriental affinities. *Am. J. Hum. Genet.* 53: 609–618.

Passarino, G., Semino, O., Pepe, G., Shrestha, S.L., Modiano, G., and Santachiara Benerecetti, A.S. 1992. MtDNA polymorphisms among Tharus of eastern Terai (Nepal). *Gene Geogr* 6: 139–147.

Pathak, A.K. 2020. Identifying admixture and genetic ancestry in human populations via genetic drift pattern. *Polymorphism* 4: 5–20.

Patnaik, R., Chauhan, P.R., Rao, M., Blackwell, B., Skinner, A., Sahni, A., Chauhan, M., and Khan, H. 2009. New geochronological, paleoclimatological, and archaeological data from the Narmada Valley hominin locality, central India. *Journal of Human Evolution* 56: 114–133.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. 2012. Ancient admixture in human history. *Genetics* 192: 1065–1093.

Peter, B.M. 2016. Admixture, Population Structure, and F-Statistics. *Genetics* 202: 1485–1501.

Petkova, D., Novembre, J., and Stephens, M. 2016. Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics* 48: 94–100.

Petraglia, M., Korisettar, R., Boivin, N., Clarkson, C., Ditchfield, P., Jones, S., Koshy, J., Lahr, M.M., Oppenheimer, C., and Pyle, D. 2007. Middle Paleolithic assemblages from the Indian subcontinent before and after the Toba super-eruption. *Science* 317: 114–116.

Petraglia, M.D., and Allchin, B. 2007. The evolution and history of human populations in South Asia: Inter-disciplinary studies in archaeology, biological anthropology, linguistics and genetics (Springer Science & Business Media).

Petraglia, M.D., Ditchfield, P., Jones, S., Korisettar, R., and Pal, J. 2012. The Toba volcanic super-eruption, environmental change, and hominin occupation history in India over the last 140,000 years. *Quaternary International* 258: 119–134.

Petraglia, M.D., Haslam, M., Fuller, D.Q., Boivin, N., and Clarkson, C. 2010. Out of Africa: new hypotheses and evidence for the dispersal of Homo sapiens along the Indian Ocean rim. *Annals of Human Biology* 37: 288–311.

Petraglia, M.D., Korisettar, R., Bai, M.K., Boivin, N.L., Janardhana, B., Clarkson, C., Cunningham, K., Ditchfield, P., Fuller, D., and Hampson, J. 2009. Human occupation, adaptation and behavioral change in the Pleistocene and Holocene of South India: Recent investigations in the Kurnool District, Andhra Pradesh. *Eurasian Prehistory* 6: 119–166.

Petrie, C.A. 2013. Petrie, C. A. in Oxford Handbook of Cities in History (ed P. Clark) 139–173 (Oxford University Press, Oxford, 2013). *Oxford University Press* 130–173.

Petrie, C.A., Singh, R.N., Bates, J., Dixit, Y., French, C.A., Hodell, D.A., Jones, P.J., Lancelotti, C., Lynam, F., and Neogi, S. 2017. Adaptation to Variable Environments, Resilience to Climate Change: Investigating Land, Water and Settlement in Indus Northwest India. *Current Anthropology* 58: 000–000.

Pickrell, J.K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., and Reich, D. 2014. Ancient west Eurasian ancestry in southern and eastern Africa. *PNAS* 111: 2632–2637.

Pickrell, J.K., and Pritchard, J.K. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics* 8: e1002967.

Pickrell, J.K., and Reich, D. 2014. Toward a new history and geography of human genes informed by ancient DNA. *Trends in Genetics* 30: 377–389.

Pinz, K.G., and Bogenhagen, D.F. 1998. Efficient repair of abasic sites in DNA by mitochondrial enzymes. *Molecular and Cellular Biology* 18: 1257–1265.

Porat, N., Chazan, M., Grün, R., Aubert, M., Eisenmann, V., and Horwitz, L.K. 2010. New radiometric ages for the Fauresmith industry from Kathu Pan, southern Africa: Implications for the Earlier to Middle Stone Age transition. *Journal of Archaeological Science* 37: 269–283.

Possehl, G.L. 1999. Possehl, G. L. Indus Age: The Beginnings. (University of Pennsylvania Museum, Philadelphia, PA, 1999).

Possehl, G.L. 2002. Possehl, G. L. The Indus Civilisation: A Contemporary Perspective. (AltaMira, Lanham, MD, 2002).

Posth, C., Wißing, C., Kitagawa, K., Pagani, L., van Holstein, L., Racimo, F., Wehrberger, K., Conard, N.J., Kind, C.J., and Bocherens, H. 2017. Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. *Nature Communications* 8: 16046.

Poznik, G.D., Henn, B.M., Yee, M.-C., Sliwerska, E., Euskirchen, G.M., Lin, A.A., Snyder, M., Quintana-Murci, L., Kidd, J.M., and Underhill, P.A. 2013. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* 341: 562–565.

Price, A.L., Tandon, A., Patterson, N., Barnes, K.C., Rafaels, N., Ruczinski, I., Beaty, T.H., Mathias, R., Reich, D., and Myers, S. 2009. Sensitive Detection of Chromosomal Segments of Distinct Ancestry in Admixed Populations. *PLOS Genetics* 5: e1000519.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Prüfer, K., Filippo, C. de, Grote, S., Mafessoni, F., Korlević, P., et al. 2017. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* 358: 655–658.

Prüfer, K., Posth, C., Yu, H., Stoessel, A., Spyrou, M.A., et al. 2021. A genome sequence from a modern human skull over 45,000 years old from Zlatý kůň in Czechia. *Nature Ecology & Evolution* 1–6.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505: 43–49.

Purana, B. 1971. Bhagavata Purana. Sanskrit text with an English Translation (Gorakhpur: Gita Press. Traducción de CL Goswami).

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., and Daly, M.J. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81.

Pyle, A., Hudson, G., Wilson, I.J., Coxhead, J., Smertenko, T., Herbert, M., Santibanez-Koref, M., and Chinnery, P.F. 2015. Extreme-depth re-sequencing of mitochondrial DNA finds no evidence of paternal transmission in humans. *PLoS Genetics* 11.

Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A., Zerjal, T., Tyler-Smith, C., and Mehdi, S.Q. 2002. Y-chromosomal DNA variation in Pakistan. *Am. J. Hum. Genet.* 70: 1107–1124.

Qin, P., and Stoneking, M. 2015. Denisovan ancestry in East Eurasian and native American populations. *Molecular Biology and Evolution* 32: 2665–2674.

Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., and Santachiara-Benerecetti, A.S. 2004. Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *The American Journal of Human Genetics* 74: 827–845.

Quintana-Murci, L., Krausz, C., Zerjal, T., Sayar, S.H., Hammer, M.F., Mehdi, S.Q., Ayub, Q., Qamar, R., Mohyuddin, A., and Radhakrishna, U. 2001. Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *The American Journal of Human Genetics* 68: 537–542.

Racimo, F., Sikora, M., Vander Linden, M., Schroeder, H., and Lalueza-Fox, C. 2020. Beyond broad strokes: sociocultural insights from the study of ancient genomes. *Nature Reviews Genetics* 21: 355–366.

Radhakrishna, B. 1999. Vedic Sarasvati and the dawn of Indian civilization. *Memoirs-Geological Society of India* 5–14.

Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., et al. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505: 87–91.

Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., et al. 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349.

Rai, N., Taher, N., Singh, M., Chaubey, G., Jha, A.N., Singh, L., and Thangaraj, K. 2014. Relic excavated in western India is probably of Georgian Queen Ketevan. *Mitochondrion* 14: 1–6.

Raj, A., Stephens, M., and Pritchard, J.K. 2014. fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics* 197: 573–589.

Rajani, M.B., and Rajawat, A.S. 2011. Potential of satellite based sensors for studying distribution of archaeological sites along palaeo channels: Harappan sites a case study. *Journal of Archaeological Science* 38: 2010–2016.

Ralph, P., and Coop, G. 2013. The Geography of Recent Genetic Ancestry across Europe. *PLOS Biology* 11: e1001555.

Rao, R.P., Yadav, N., Vahia, M.N., Joglekar, H., Adhikari, R., and Mahadevan, I. 2009. Entropic evidence for linguistic structure in the Indus script. *Science* 324: 1165–1165.

Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., et al. 2011. An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334: 94–98.

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463: 757–762.

Rau, F., and Sidwell, P. 2019. The Munda Maritime Hypothesis. *Journal of the Southeast Asian Linguistics Society* 12: 35–57.

Raza, A., Firasat, S., Khaliq, S., Abid, A., Shah, S.S., Mehdi, S.Q., and Mohyuddin, A. 2013. HLA class I and II polymorphisms in the Gujjar population from Pakistan. *Immunological Investigations*.

Reich, D. 2018. Who we are and how we got here: Ancient DNA and the new science of the human past (Oxford University Press).

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468: 1053–1060.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., et al. 2012. Reconstructing Native American population history. *Nature* 488: 370–374.

Reich, D., Patterson, N., Jager, P.L.D., McDonald, G.J., Waliszewska, A., et al. 2005. A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nat Genet* 37: 1113–1118.

Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M.R., et al. 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* 89: 516–528.

Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. 2009. Reconstructing Indian population history. *Nature* 461.

Renfrew, C. 1996. Language families and the spread of farming. *The Origins and Spread of Agriculture and Pastoralism in Eurasia*.

Renfrew, C. 1991. The coming of the Aryans to Iran and India and the cultural and ethnic identity of the Dāsas. By Asko Parpola. (Studia Orientalia, Vol. 64.) pp. 195–302, 33 figs. Helsinki, The Finnish Oriental Society, 1988. *Journal of the Royal Asiatic Society* 1: 106–109.

Richards, M., Bandelt, H.-J., Kivisild, T., and Oppenheimer, S. 2006. A model for the dispersal of modern humans out of Africa. In Human Mitochondrial DNA and the Evolution of Homo Sapiens, (Springer), pp. 225–265.

Richards, M., Corte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H.-J., and Sykes, B. 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. *American Journal of Human Genetics* 59: 185.

Richter, D., Grün, R., Joannes-Boyau, R., Steele, T.E., Amani, F., et al. 2017. The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. *Nature* 546: 293–296.

Ringbauer, H., Coop, G., and Barton, N.H. 2017. Inferring Recent Demography from Isolation by Distance of Long Shared Sequence Blocks. *Genetics* 205: 1335–1351.

Robbins Schug, G., Blevins, K.E., Cox, B., Gray, K., and Mushrif-Tripathy, V. 2013. Infection, disease, and biosocial processes at the end of the Indus Civilization. *PLoS ONE* 8: e84814.

Rogers, A.R., and Jorde, L.B. 1996. Ascertainment bias in estimates of average heterozygosity. *American Journal of Human Genetics* 58: 1033.

Roland, J.G. 2018. Jewish Communities of India: Identity in a Colonial Era (Routledge).

Romero, I.G., Manica, A., Goudet, J., Handley, L.L., and Balloux, F. 2009. How accurate is the current picture of human genetic variation? *Heredity (Edinb)* 102: 120–126.

Rootsi, S., Behar, D.M., Järve, M., Lin, A.A., Myres, N.M., Passarelli, B., Poznik, G.D., Tzur, S., Sahakyan, H., and Pathak, A.K. 2013. Phylogenetic applications of whole Y-chromosome sequences and the Near Eastern origin of Ashkenazi Levites. *Nature Communications* 4: 2928.

Rosenberg, N.A., and Feldman, M.W. 2002. The relationship between coalescence times and population divergence times. *Modern Developments in Theoretical Population Genetics* 130: 164.

Rosenberg, N.A., Mahajan, S., Gonzalez-Quevedo, C., Blum, M.G., Nino-Rosales, L., Ninis, V., Das, P., Hegde, M., Molinari, L., and Zapata, G. 2006. Low levels of genetic divergence across geographically and linguistically diverse populations from India. *PLoS Genetics* 2: e215.

Rossi, M.N., Carbone, M., Mostocotto, C., Mancone, C., Tripodi, M., Maione, R., and Amati, P. 2009. Mitochondrial localization of PARP-1 requires interaction with mitofilin and is involved in the maintenance of mitochondrial DNA integrity. *Journal of Biological Chemistry* 284: 31616–31624.

Roy, K. 2012. Hinduism and the Ethics of Warfare in South Asia: From Antiquity to the Present (Cambridge University Press).
Roy, T., Unisa, S., and Bhatt, M. 2001. Growth of Parsi Population in India p.
Roychoudhury, S., Roy, S., Basu, A., Banerjee, R., Vishwanathan, H., Usha Rani, M.V., Sil, S.K., Mitra, M., and Majumder, P.P. 2001. Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum. Genet.* 109: 339–350.
Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., and Page, D.C. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423: 873–876.
Sahakyan, H., Kashani, B.H., Tamang, R., Kushniarevich, A., Francis, A., et al. 2017. Origin and spread of human mitochondrial DNA haplogroup U7. *Scientific Reports* 7: 46044.
Sahle, Y., Morgan, L.E., Braun, D.R., Atnafu, B., and Hutchings, W.K. 2014. Chronological and behavioral contexts of the earliest Middle Stone Age in the Gademotta formation, main Ethiopian rift. *Quaternary International* 331: 6–19.
Sahoo, S., Singh, A., Himabindu, G., Banerjee, J., Sitalaximi, T., Gaikwad, S., Trivedi, R., Endicott, P., Kivisild, T., and Metspalu, M. 2006. A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc Natl Acad Sci U S A* 103.
Sankararaman, S., Mallick, S., Patterson, N., and Reich, D. 2016. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. *Current Biology* 26: 1241–1247.
Sankararaman, S., Patterson, N., Li, H., Pääbo, S., and Reich, D. 2012. The Date of Interbreeding between Neandertals and Modern Humans. *PLOS Genetics* 8: e1002947.
Sankararaman, S., Sridhar, S., Kimmel, G., and Halperin, E. 2008. Estimating Local Ancestry in Admixed Populations. *The American Journal of Human Genetics* 82: 290–303.
Sankhyan, A. 1997a. Fossil clavicle of a middle Pleistocene hominid from the central Narmada Valley, India. *Journal of Human Evolution* 32: 3–16.
Sankhyan, A. 2005. New fossils of early stone age man from central Narmada valley. *Current Science* 88: 704–707.
Sankhyan, A.R. 1997b. A new human fossil find from the Central Narmada basin and its chronology. *Current Science* 73: 1110–1111.
Sankhyan, A.R., Badam, G.L., Dewangan, L.N., Chakraborty, S., Prabha, S., Kundu, S., and Chakravarty, R. 2012. New postcranial hominin fossils from the Central Narmada Valley, India. *Advances in Anthropology* 2: 125.
Saraswat, K. 1993. Plant economy of late Harappans at Hulas. *Purattatva* 23: 1–12.
Sarkar, A., Mukherjee, A.D., Bera, M.K., Das, B., Juyal, N., Morthekai, P., Deshpande, R.D., Shinde, V.S., and Rao, L.S. 2016. Oxygen isotope in archaeological bioapatites from India: Implications to climate change and decline of Bronze Age Harappan civilization. *Scientific Reports* 6: 26555.
Scerri, E.M.L., Thomas, M.G., Manica, A., Gunz, P., Stock, J.T., et al. 2018. Did Our Species Evolve in Subdivided Populations across Africa, and Why Does It Matter? *Trends in Ecology & Evolution* 33: 582–594.
Scheib, C.L., Li, H., Desai, T., Link, V., Kendall, C., et al. 2018. Ancient human parallel lineages within North America contributed to a coastal expansion. *Science* 360: 1024–1027.
Schimmel, A. 1982. Islam in India and Pakistan (Brill).

Schlebusch, C.M., Malmström, H., Günther, T., Sjödin, P., Coutinho, A., et al. 2017. Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* 358: 652–655.

Schraiber, J.G., and Akey, J.M. 2015. Methods and models for unravelling human evolutionary history. *Nature Reviews Genetics* 16: 727–740.

Schwartz, M., and Vissing, J. 2004. No evidence for paternal inheritance of mtDNA in patients with sporadic mtDNA mutations. *Journal of the Neurological Sciences* 218: 99–101.

Schwartz, M., and Vissing, J. 2002. Paternal inheritance of mitochondrial DNA. *New England Journal of Medicine* 347: 576–580.

Seguin-Orlando, A., Korneliussen, T.S., Sikora, M., Malaspinas, A.-S., Manica, A., et al. 2014. Genomic structure in Europeans dating back at least 36,200 years. *Science* 346: 1113–1118.

Sekar, C.C. 1948. Some aspects of Parsi demography. *Human Biology* 20: 47–89.

Sekido, R., and Lovell-Badge, R. 2008. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* 453: 930–934.

Sengupta, D., Choudhury, A., Basu, A., and Ramsay, M. 2016. Population stratification and underrepresentation of Indian subcontinent genetic diversity in the 1000 genomes project dataset. *Genome Biol Evol* 8.

Sengupta, S., Zhivotovsky, L.A., King, R., Mehdi, S.Q., Edmonds, C.A., Chow, C.E., Lin, A.A., Mitra, M., Sil, S.K., and Ramesh, A. 2006. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central Asian pastoralists. *Am J Hum Genet* 78.

Shah, A.M., Tamang, R., Moorjani, P., Rani, D.S., Govindaraj, P., Kulkarni, G., Bhattacharya, T., Mustak, M.S., Bhaskar, L.V., and Reddy, A.G. 2011. Indian Siddis: African descendants with Indian admixture. *Am J Hum Genet* 89.

Sharma, A.K. 1999. The departed Harappans of Kalibangan (Sundeep Prakashan).

Sharma, D.V., Nauriyal, K.C., and Prabhakar, V.N. 2007. Excavations at Sanauli 2005–06: A Harappan Necropolis in the Upper Ganga-Yamuna Doab. *Puratattva* 36: 166–179.

Sharma, G., Tamang, R., Chaudhary, R., Singh, V.K., Shah, A.M., Anugula, S., Rani, D.S., Reddy, A.G., Eaaswarkhanth, M., and Chaubey, G. 2012. Genetic affinities of the central Indian tribal populations. *PloS One* 7: e32546.

Sharon, G. 2007. Acheulian large flake industries: technology, chronology, and significance (British Archaeological Reports Ltd).

Shinde, V. 2016. Harappan Civilization: Current Perspective and its Contribution–By Dr. Vasant Shinde. *John Wiley & Sons* Vol 31:

Shinde, V., Deshpande, S., Osada, T., and Uno, T. 2006. Basic Issues in Harappan Archaeology: Some Thoughts. *Ancient Asia* 1.

Shinde, V., Narasimhan, V.M., Rohland, N., Mallick, S., Mah, M., et al. 2019. An Ancient Harappan Genome Lacks Ancestry from Steppe Pastoralists or Iranian Farmers. *Cell* 179: 729-735.e10.

Shinde, V., Osada, T., and Kumar, M. 2011. Excavations at Farmana, District Rohtak, Haryana, India, 2006–2008 (Indus Project, Research Institute for Humanity and Nature).

Shinde, V.S., Kim, Y.J., Woo, E.J., Jadhav, N., Waghmare, P., et al. 2018. Archaeological and anthropological studies on the Harappan cemetery of Rakhigarhi, India. *PLOS ONE* 13: e0192299.

Shokolenko, I., Venediktova, N., Bochkareva, A., Wilson, G.L., and Alexeyev, M.F. 2009. Oxidative stress induces degradation of mitochondrial DNA. *Nucleic Acids Res* 37: 2539–2548.

Shroff, Z., and Castro, M.C. 2011. The potential impact of intermarriage on the population decline of the Parsis of Mumbai, India. *Demographic Research* 25: 545–564.

Sikora, M., Carpenter, M.L., Moreno-Estrada, A., Henn, B.M., Underhill, P.A., Sánchez-Quinto, F., Zara, I., Pitzalis, M., Sidore, C., and Busonero, F. 2014. Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe. *PLoS Genet* 10: e1004353.

Silva, M., Oliveira, M., Vieira, D., Brandao, A., Rito, T., Pereira, J.B., Fraser, R.M., Hudson, B., Gandini, F., and Edwards, C. 2017. A genetic chronology for the Indian subcontinent points to heavily sex-biased dispersals. *BMC Evol Biol* 17.

Singh, A., Thomsen, K.J., Sinha, R., Buylaert, J.-P., Carter, A., et al. 2017. Counter-intuitive influence of Himalayan river morphodynamics on Indus Civilisation urban settlements. *Nature Communications* 8: 1617.

Singh, G. 1971. The Indus valley culture. *Archaeology in Oceania* 6: 177–189.

Singh, S., Singh, A., Rajkumar, R., Kumar, K.S., Samy, S.K., Nizamuddin, S., Singh, A., Sheikh, S.A., Peddada, V., and Khanna, V. 2016. Dissecting the influence of Neolithic demic diffusion on Indian Y-chromosome pool through J2-M172 haplogroup. *Scientific Reports* 6.

Sitaramiah, V., and Sītārāmayya, V. 1972. Valmiki Ramayana (New Delhi: Sahitya Akademi).

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., and Bieri, T. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423: 825–837.

Skoglund, P., and Jakobsson, M. 2011. Archaic human ancestry in East Asia. *PNAS* 108: 18301–18306.

Skoglund, P., Malmström, H., Omrak, A., Raghavan, M., Valdiosera, C., et al. 2014. Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* 344: 747–750.

Skoglund, P., and Mathieson, I. 2018. Ancient Genomics of Modern Humans: The First Decade. *Annual Review of Genomics and Human Genetics* 19: 381–404.

Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., et al. 2016. Genomic insights into the peopling of the Southwest Pacific. *Nature* 538: 510–513.

Skoglund, P., Thompson, J.C., Prendergast, M.E., Mittnik, A., Sirak, K., et al. 2017. Reconstructing Prehistoric African Population Structure. *Cell* 171: 59-71.e21.

Skov, L., Coll Macià, M., Sveinbjörnsson, G., Mafessoni, F., Lucotte, E.A., et al. 2020. The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature* 582: 78–83.

Slapak, O., and Yiśra'el, M. 1995. The Jews of India: a story of three communities/edited by Orpa Slapak; [translations from Hebrew, David Louvish] (Israel Museum).

Slatkin, M., and Racimo, F. 2016. Ancient DNA and human history. *Proceedings of the National Academy of Sciences* 113: 6380–6387.

Slavitt, D.R. 2015. The Rig Veda: First Mandala (Anaphora Literary Press).

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., and Richards, M.B. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *The American Journal of Human Genetics* 84: 740–759.

Sonakia, A., and Kennedy, K.A. 1985. Skull cap of an early man from the Narmada valley alluvium (Pleistocene) of central India. *American Anthropologist* 87: 612–616.
Srinivasan, L. 2000. The Land and People Kuru. *Geographical Review of India* 62: 39–47.
Stein, B., and Arnold, D. 2010. A History of India (New York, NY: John Wiley & Sons).
Stoneking, M., and Krause, J. 2011. Learning about human population history from ancient and modern genomes. *Nature Reviews Genetics* 12: 603.
Stringer, C. 2002. Modern human origins: progress and prospects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 357: 563–579.
Stringer, C. 2016. The origin and evolution of Homo sapiens. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371: 20150237.
Sundquist, A., Fratkin, E., Do, C.B., and Batzoglou, S. 2008. Effect of genetic divergence in identifying ancestral origin using HAPAA. *Genome Res.* 18: 676–682.
Szczesny, B., Tann, A.W., Longley, M.J., Copeland, W.C., and Mitra, S. 2008. Long patch base excision repair in mammalian mitochondrial genomes. *Journal of Biological Chemistry* 283: 26349–26356.
Tamang, R., Chaubey, G., Nandan, A., Govindaraj, P., Singh, V.K., et al. 2018. Reconstructing the demographic history of the Himalayan and adjoining populations. *Hum Genet* 137: 129–139.
Tamang, R., and Thangaraj, K. 2012. Genomic view on the peopling of India. *Investig Genet* 3: 20.
Tambets, K., Yunusbayev, B., Hudjashov, G., Ilumäe, A.-M., Rootsi, S., et al. 2018. Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. *Genome Biology* 19: 139.
Tang, H., Choudhry, S., Mei, R., Morgan, M., Rodriguez-Cintron, W., Burchard, E.G., and Risch, N.J. 2007. Recent Genetic Selection in the Ancestral Admixture of Puerto Ricans. *The American Journal of Human Genetics* 81: 626–633.
Tang, H., Coram, M., Wang, P., Zhu, X., and Risch, N. 2006. Reconstructing Genetic Ancestry Blocks in Admixed Individuals. *The American Journal of Human Genetics* 79: 1–12.
Tang, H., Peng, J., Wang, P., and Risch, N.J. 2005. Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology* 28: 289–301.
Tätte, K., Pagani, L., Pathak, A.K., Kõks, S., Duy, B.H., et al. 2019. The genetic legacy of continental scale admixture in Indian Austroasiatic speakers. *Scientific Reports* 9: 3818.
Taylor, R.W., McDonnell, M.T., Blakely, E.L., Chinnery, P.F., Taylor, G.A., Howell, N., Zeviani, M., Briem, E., Carrara, F., and Turnbull, D.M. 2003. Genotypes from patients indicate no paternal mitochondrial DNA contribution. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 54: 521–524.
Teixeira, J.C., and Cooper, A. 2019. A 'Denisovan'genetic history of recent human evolution (PeerJ Preprints).
Terrenato, L., Shrestha, S., Dixit, K.A., Luzzatto, L., Modiano, G., Morpurgo, G., and Arese, P. 1988. Decreased malaria morbidity in the Tharu people compared to sympatric populations in Nepal. *Ann Trop Med Parasitol* 82: 1–11.
Thangaraj, K., Chaubey, G., Kivisild, T., Rani, D.S., Singh, V.K., Ismail, T., Carvalho-Silva, D., Metspalu, M., Bhaskar, L., and Reddy, A.G. 2008. Maternal footprints of Southeast Asians in North India. *Human Heredity* 66: 1–9.

Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A.G., Singh, V.K., Rasalkar, A.A., and Singh, L. 2005. Reconstructing the origin of Andaman Islanders. *Science* 308: 996.

Thangaraj, K., Chaubey, G., Singh, V.K., Vanniarajan, A., Thanseem, I., Reddy, A.G., and Singh, L. 2006. In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. *BMC Genomics* 7:.

Thanseem, I., Thangaraj, K., Chaubey, G., Singh, V.K., Bhaskar, L.V., Reddy, B.M., Reddy, A.G., and Singh, L. 2006. Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC Genetics* 7: 42.

The International HapMap Consortium, Frazer (Principal Investigator), K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., et al. 2009. The Genetic Structure and History of Africans and African Americans. *Science* 324: 1035–1044.

Torroni, A., Achilli, A., Macaulay, V., Richards, M., and Bandelt, H.-J. 2006. Harvesting the fruit of the human mtDNA tree. *TRENDS in Genetics* 22: 339–345.

Torroni, A., Bandelt, H.-J., D'urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M.-L., and Bonné-Tamir, B. 1998. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *The American Journal of Human Genetics* 62: 1137–1152.

Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M., and Wallace, D.C. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* 53: 563–590.

Tripathi, J.K., Bock, B., Rajamani, V., and Eisenhauer, A. 2004. Is Rhiver Ghaggar, Saraswati? Geochemical constraints. *Current Science* 87: 1141–1144.

Tryon, C. 2006. "Early" Middle Stone Age Lithic Technology of the Kapthurin Formation (Kenya). *Current Anthropology* 47: 367–375.

Tryon, C.A. 2003. The Acheulian to Middle Stone Age transition: Tephrostratigraphic context for archaeological change in the Kapthurin Formation, Kenya.

Tuppen, H.A., Blakely, E.L., Turnbull, D.M., and Taylor, R.W. 2010. Mitochondrial DNA mutations and human disease. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1797: 113–128.

Underhill, P.A., Poznik, G.D., Rootsi, S., Järve, M., Lin, A.A., Wang, J., Passarelli, B., Kanbar, J., Myres, N.M., and King, R.J. 2015. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *European Journal of Human Genetics* 23: 124–131.

Undevia, J. 1973. Population genetics of the Parsis: Comparison of genetical characteristics of the present Parsi population with its ancestral and affiliated groups (Field research projects).

UNESCOParzor 2019. Parzor | The UNESCO Parsi Zoroastrian Project.

UNSD, E. 1999. Standard Country or Area Codes for Statistics Use. M49.

Ṿail, S. 2002. India's Jewish heritage: ritual, art & life-cycle/edited by Shalva Weil (Marg Publications).

Valdiya, K.S. 2017. Prehistoric River Saraswati, Western India: Geological Appraisal and Social Aspects (Cham: Springer International Publishing: Imprint: Springer).

Van Driem, G. 2001. Languages of the Himalayas: an ethnolinguistic handbook of the greater Himalayan region (Brill).

Van Oven, M., and Kayser, M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation* 30: E386–E394.
Veeramah, K.R., and Hammer, M.F. 2014. The impact of whole-genome sequencing on the reconstruction of human population history. *Nature Reviews Genetics* 15: 149–162.
Verdu, P., Pemberton, T.J., Laurent, R., Kemp, B.M., Gonzalez-Oliver, A., et al. 2014. Patterns of Admixture and Population Structure in Native Populations of Northwest North America. *PLOS Genetics* 10: e1004530.
Vernot, B., and Pääbo, S. 2018. The Predecessors Within . . . *Cell* 173: 6–7.
Wakeley, J. 1996. Pairwise differences under a general model of population subdivision. *Journal of Genetics* 75: 81–89.
Waldman, Y.Y., Biddanda, A., Davidson, N.R., Billing-Ross, P., Dubrovsky, M., et al. 2016a. The Genetics of Bene Israel from India Reveals Both Substantial Jewish and Indian Ancestry. *PLOS ONE* 11: e0152056.
Waldman, Y.Y., Biddanda, A., Dubrovsky, M., Campbell, C.L., Oddoux, C., Friedman, E., Atzmon, G., Halperin, E., Ostrer, H., and Keinan, A. 2016b. The genetic history of Cochin Jews from India. *Hum Genet* 135: 1127–1143.
Wall, J.D., Stawiski, E.W., Ratan, A., Kim, H.L., Kim, C., et al. 2019. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* 576: 106–111.
Wang, C.-C., Yeh, H.-Y., Popov, A.N., Zhang, H.-Q., Matsumura, H., et al. 2020. The Genomic Formation of Human Populations in East Asia. *BioRxiv* 2020.03.25.004606.
Wang, H.-W., Li, Y.-C., Sun, F., Zhao, M., Mitra, B., Chaudhuri, T.K., Regmi, P., Wu, S.-F., Kong, Q.-P., and Zhang, Y.-P. 2012. Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. *Journal of Human Genetics* 57: 228.
Wang, S., Lewis, C.M., Jakobsson, M., Ramachandran, S., Ray, N., et al. 2007. Genetic variation and population structure in native Americans. *PLoS Genet.* 3: e185.
Wangkumhang, P., and Hellenthal, G. 2018. Statistical methods for detecting admixture. *Curr. Opin. Genet. Dev.* 53: 121–127.
Watkins, W., Thara, R., Mowry, B., Zhang, Y., Witherspoon, D., Tolpinrud, W., Bamshad, M., Tirupati, S., Padmavati, R., and Smith, H. 2008. Genetic variation in South Indian castes: evidence from Y-chromosome, mitochondrial, and autosomal polymorphisms. *BMC Genetics* 9: 86.
Wedage, O., Amano, N., Langley, M.C., Douka, K., Blinkhorn, J., et al. 2019a. Specialized rainforest hunting by Homo sapiens ~45,000 years ago. *Nat Commun* 10: 1–8.
Wedage, O., Picin, A., Blinkhorn, J., Douka, K., Deraniyagala, S., et al. 2019b. Microliths in the South Asian rainforest ~45–4 ka: New insights from Fa-Hien Lena Cave, Sri Lanka. *PLOS ONE* 14: e0222606.
Weir, B.S., and Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 1358–1370.
Wells, R.S., Yuldasheva, N., Ruzibakiev, R., Underhill, P.A., Evseeva, I., et al. 2001. The Eurasian Heartland: A continental perspective on Y-chromosome diversity. *PNAS* 98: 10244–10249.
Westaway, K.E., Louys, J., Awe, R.D., Morwood, M.J., Price, G.J., Zhao, J., Aubert, M., Joannes-Boyau, R., Smith, T., and Skinner, M.M. 2017. An early modern human presence in Sumatra 73,000–63,000 years ago. *Nature* 548: 322.

White, T.D., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G.D., Suwa, G., and Howell, F.C. 2003. Pleistocene homo sapiens from middle awash, ethiopia. *Nature* 423: 742.

Witzel, E.M. 2012. The origins of the world's mythologies (Oxford University Press).

Witzel, M. 2016. Early Sanskritization. Origins and Development of the Kuru State. *Electronic Journal of Vedic Studies* Vol 1: Year: 2014 .

Witzel, M. 2003. Ein Fremdling im Rgveda: Indo-Aryan migration debate. *Journal of Indo-European Studies* 31: 107–185.

Wright, R.P. 2010. The ancient Indus: urbanism, economy, and society.

Xing, J., Watkins, W.S., Hu, Y., Huff, C.D., Sabo, A., Muzny, D.M., Bamshad, M.J., Gibbs, R.A., Jorde, L.B., and Yu, F. 2010. Genetic diversity in India and the inference of Eurasian population expansion. *Genome Biology* 11: R113.

Xu, D., Pavlidis, P., Taskent, R.O., Alachiotis, N., Flanagan, C., DeGiorgio, M., Blekhman, R., Ruhl, S., and Gokcumen, O. 2017. Archaic Hominin Introgression in Africa Contributes to Functional Salivary MUC7 Genetic Variation. *Mol Biol Evol* 34: 2704–2715.

Yang, J.J., Cheng, C., Devidas, M., Cao, X., Fan, Y., et al. 2011. Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. *Nat Genet* 43: 237–241.

Yang, M.A., Fan, X., Sun, B., Chen, C., Lang, J., Ko, Y.-C., Tsang, C., Chiu, H., Wang, T., and Bao, Q. 2020. Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science*.

Yardumian, A., and Schurr, T.G. 2019. The Geography of Jewish Ethnogenesis. *Journal of Anthropological Research* 75: 206–234.

Yelmen, B., Mondal, M., Marnetto, D., Pathak, A.K., Montinaro, F., Gallego Romero, I., Kivisild, T., Metspalu, M., and Pagani, L. Ancestry-Specific Analyses Reveal Differential Demographic Histories and Opposite Selective Pressures in Modern South Asian Populations. *Mol Biol Evol*.

Yunusbayev, B., Metspalu, M., Järve, M., Kutuev, I., Rootsi, S., et al. 2012. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol. Biol. Evol.* 29: 359–365.

Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litvinov, S., et al. 2015. The Genetic Legacy of the Expansion of Turkic-Speaking Nomads across Eurasia. *PLOS Genetics* 11: e1005068.

ACKNOWLEDGEMENTS

As the saying goes, “It’s never too late to do something”. My dissertation journey became lengthier, to a certain degree, than I had anticipated. Nevertheless, as happens with every long journey, however tough it has been, I have had my share of good memories, experiences and a positive learning curve. Start of this journey, I had to witness the unfortunate, accidental demise of my beloved Dad, an academician himself, who would have been the happiest person to see me complete this pursuit. This accident had shattered me so badly that it took years to get out of one of the deepest pains I ever had. I would ever remain indebted for the support extended by my supervisors, colleagues, family and friends which kept me afloat and helped me sail through this rough phase and held me on the right track.

First and foremost, I would like to show my deepest gratitude to Professor Richard Villems for giving me this opportunity to work under him. His cherished guidance, constructive inputs and abundance of information always helped me like a light at the end of tunnel. I am extremely thankful to Professor Gyaneshwer for his helps during the PhD application and making me feel at home when I moved to Tartu; his consistent feedback and guidance have been instrumental.

I am extremely grateful to all the sample donors who trusted us and gave their consent to use their samples in our scientific investigations. My gratitude goes out to all my co-authors included in this thesis who made this possible.

I would like to extend my sincere gratitude to Professor Toomas Kivisild, who has always been a source of inspiration, wisdom and ready to lend helping hand through his prompt email exchanges. I am enormously thankful to Professor Mait for his easy accessibility, openness and problem-solving approaches that along-with his criticial assessment skills guided me through difficulties, countless times. Professor Kristiina, I am grateful for your positive vibes and optimistic words – “I keep asking to let you know, I have your back” those have been super helpful to keep my spirits up.

I am greatly appreciative of the EBC - now part of the Institute of Genomics, and Institute of Molecular and Cell Biology, for their positive and friendly work culture that makes newbies comfortable and motivated. Estonia has become second home for me, and the credit mostly goes to our colleagues and friends. I am grateful to Ene, Siiri and Maere for their support and suggestions in getting me acquainted with the lab and its environment. I would like to express my gratitude to Jüri for broadening my knowledge about wine and the Estonian weather and helping me with procuring reagents for lab experiments. I also thank Tuuli and Anu for their technical assistances, and Merilin and Maritza for their help with administrative issues.

I am highly obliged to Lena, who has consistently been a source of motivation and confidence booster with her amazing smile and cheerful persona, apart from her valuable technical help. I would like to express my gratitude to Luca for his positivity, critical inputs and constructive advices. I am deeply thankful to

Francesco, who always greeted me with his evergreen smile and inspirational word – *Bravo Bravo!* Your consistent backup, technical helps, and suggestions have been instrumental. Mayukh, Linda, Davide and Bayazit thanks for being there, whenever I needed and helping me out with your analytical skills. I would like to thank Anne-Mai, Lehti, and Helen for their constant support and willingness to assist.

Rodrigo and Hovik- my moral boosting partners with whom I have been sharing not just the "pigeon room" but, even rooms during conferences and summer days, it has been pleasure talking to you guys endlessly on a range of issues – from technical to general. I wish to thank Chandana for being there in the initial phase of my journey and the encouragement extended. I am also thankful to Erwan and Monika for setting example of persistence and patience, their useful suggestions and encouraging remarks "keep going" – helped a lot.

I wish to extend my thanks to Mari, for her kind help in writing Estonian summary, and Georgi – the first ones I saw defending. My gratitude list seems to be long. I take this furthermore by extending it to, Freddi, Vasili, Ludovica, Erika, Marcel, Meriam for their supportive spirit and cheering words which made my lab time memorable. I would also like to thank Kai, Katri and Tina, Burak, Stefania, and Mathilde for a pleasant and cohesive environment.

I express my gratitude to Doron for his fun filled conversations and his encouraging behavior. I extend my gratitude to Philip for several interesting conversations and suggestions.

I am extremely thankful to Professor Maido Remm for kindly agreeing to be my internal opponent and helping me sail through the completion of the thesis.

This journey would be incomplete if I don't mention my family and friends back in India. I would like to thank everyone from the bottom of my heart, my siblings, in-laws, my friends and most importantly, my sweet mom, who have been a source of strength all along.

I am grateful to European Social Fund and Archimedes Foundation for DoRa scholarship for my PhD study. I am also thankful to EBC, Institute of Genomics for financial assistance towards my participations in related courses and conferences that have been quite useful in gaining skills and experiences.

Last but not the least, Katyayani and Akshith, you both have been by my side in the stressful periods ensuring that everything went by as normal as it could and bearing the brunt instead and Akshith's complaint still ringing in my ears that, "I am not giving enough time to him". I accept that without you two, it was impossible to sail along this journey. Bravo!

PUBLICATIONS

CURRICULUM VITAE

Name: Ajai Kumar Pathak
Date of Birth: January 1, 1979
Contacts: University of Tartu, Institute of Genomics, Riia 23b, 51010 Tartu, Estonia +372 737 5050
Email: ajai.kumar.pathak@ut.ee

Education:

2011–2019	Doctoral studies, University of Tartu, Faculty of Science and Technology, Institute of Molecular and Cell Biology
2002–2004	Master of Science (*MSc*) in Biotechnology, VBS Purvanchal University, Jaunpur, India
2004	*MSc* Dissertation from Centre for Cellular and Molecular Biology (CCMB), Hyderabad, India. Title: "Role of SULT1E1 Gene Polymorphism and Non-Random X-Chromosome Inactivation in Recurrent Early Pregnancy Loss (REPL)" under supervision of Dr. Lakshmi Rao and Professor Lalji Singh
2003	*MSc* Summer Training from CCMB, Hyderabad, India. Title: "Role of the Follistatin gene mutations in Polycystic Ovarian Syndrome (PCOS)" in the supervision of Dr. Lakshmi Rao and Professor Lalji Singh (May 2003-July 2003)
1995–1998	Bachelor of Science (*BSc*), Purvanchal University, Jaunpur, India

Professional Employment:

2018–...	Junior Research Fellow of Population Genetics at Estonian Biocentre, Institute of Genomics of University of Tartu, Estonia
2012–2017	Visiting Researcher at Estonian Biocentre, Estonia
2006–2011	Scientist at DS Genomic Solutions Pvt Ltd, India (Research & IP)
2004–2006	Project Assistant at Centre for Cellular and Molecular Biology (CCMB), India

Distinctions and Awards:

2019	EMBO | EMBL Symposia Travel Grant
2019	American Society for Human Genetics (ASHG) Trainee Paper Spotlight award
2013	Dora T8 short term mobility scholarship
2011	European Social Fund's Doctoral Studies and Internationalization Programme "DoRa" scholarship grant

Additional courses:

2018 Elixir popgen practical course "Population Genomics: background and tools" from 21–27 April 2018 at the CNR Institute of Genetics and Biophysics, Napoli, Italy

2013 Otto Warburg International Summer School and Research Symposium on Next Generation Sequencing and Its Impact on Genetics, August 19–26. 2013 at Max Planck Institute for Molecular Genetics, Berlin, Germany

2007 PG Diploma in Patent Laws from NALSAR University of Law, Hyderabad, India

2006 Primer course in Intellectual Property Right, World Intellectual Property Organization (WIPO)

Creative work:

2019 Organized the International Yoga Day, Tartu, in association with Indian Embassy of Finland and Estonia

2018 Scientific outreach talk on population genetics for Master Students, Department of Biotechnology, VBS Purvanchal University, Jaunpur, India

Memberships:

Member of American Society of Human Genetics (ASHG)

Life-time Member of Indian Society of Human Genetics (ISHG)

Publications:

Eaaswarkhanth, M; ***Pathak, Ajai K**; Ongaro, L; Montinaro, F; Hebbar, P; Alsmadi, O; Metspalu, M; Al-Mulla, F; Thanaraj, TA (2021). **Unraveling a fine-scale high genetic heterogeneity and recent continental connections of an Arabian Peninsula population.** European Journal of Human Genetics, 1–13, https://doi.org/10.1038/s41431-021-00861-6.

Pathak, Ajai, K (2020). **Identifying admixture and genetic ancestry in human populations via genetic drift pattern.** Polymorphism, 4, 5–20.

Yelmen, Burak; Mondal, Mayukh; Marnetto, Davide; **Pathak, Ajai K**; Montinaro, Francesco; Romero, Irene Gallego; Kivisild, Toomas; Metspalu, Mait; Pagani, Luca (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.

Pathak, Ajai K; Srivastava, Anshika; Singh, Prajjval Pratap; Das, Debashruti; Bandopadhyay, Audditiya; Singh, Pooja; Chaubey, Gyaneshwer (2019). **Historic migration to South Asia in the last two millennia: A case of Jewish and Parsi populations.** Journal of Biosciences, 44 (3), UNSP 72. https://doi.org/10.1007/s12038-019-9873-7.

Tätte, Kai; Pagani, Luca; **Pathak, Ajai K**; Kõks, Sulev; Ho Duy, Binh; Ho, Xuan Dung; Sultana, Gazi Nurun Nahar; Sharif, Mohd Istiaq; Asaduzzaman, Md; Behar, Doron M; Hadid, Yarin; Villems, Richard; Chaubey, Gyaneshwer;

Kivisild, Toomas; Metspalu, Mait (2019). **The genetic legacy of continental scale admixture in Indian Austroasiatic speakers.** Scientific Reports, 9 (1), 3818–3818. https://doi.org/10.1038/s41598-019-40399-8.

Järve, Mari; Saag, Lehti; Scheib, Christiana Lyn; **Pathak, Ajai K**; Montinaro, Francesco; Pagani, Luca; Flores, Rodrigo; Guellil, Meriam; Saag, Lauri; Tambets, Kristiina; Kushniarevich, Alena; Solnik, Anu; Varul, Liivi; Zadnikov, Stanislav; Petrauskas, Oleg; Avramenko, Maryana; Magomedov, Boris; Didenko, Serghii; Toshev, Gennadi; Bruyako, Igor ... Villems, Richard (2019). **Shifts in the Genetic Landscape of the Western Eurasian Steppe Associated with the Beginning and End of the Scythian Dominance.** Current Biology, 29 (14), 2430–2452. https://doi.org/10.1016/j.cub.2019.06.019.

Pathak, Ajai K; Kadian, Anurag; Kushniarevich, Alena; Montinaro, Francesco; Mondal, Mayukh; Ongaro, Linda; Singh, Manvendra; Kumar, Pramod; Rai, Niraj; Parik, Jüri; Metspalu, Ene; Rootsi, Siiri; Pagani, Luca; Kivisild, Toomas; Metspalu, Mait; Chaubey, Gyaneshwer; Villems, Richard (2018). **The Genetic Ancestry of Modern Indus Valley Populations from Northwest India.** American Journal of Human Genetics, 103 (6), 918–929. https://doi.org/10.1016/j.ajhg.2018.10.022.

Sahakyan, H; Kashani, BH; Tamang, R; Kushniarevich, A; Francis, A; Costa, MD; **Pathak, AK**; …Parik, J; …Metspalu, E; Pennarun, E; Karmin, M; Tamm, E; Tambets, K; …Reisberg, T; Reidla, M; Achilli, A; Olivieri, A; … Litvinov, S; Bermisheva, M; Khusnutdinova, E ... Villems, R. (2017). **Origin and spread of human mitochondrial DNA haplogroup U7**. Scientific Reports, 7, Article number: 46044. https://doi.org/10.1038/srep46044.

Chaubey, Gyaneshwer; Ayub, Qasim; Rai, Niraj; Prakash, Satya; Mushrif-Tripathy, Veena; Mezzavilla, Massimo; **Pathak, Ajai Kumar**; Tamang, Rakesh; Firasat, Sadaf; Reidla, Maere; Karmin, Monika; Rani, Deepa Selvi; Reddy, Alla G; Parik, Jüri; Metspalu, Ene; Rootsi, Siiri; Dalal, Kurush; Khaliq, Shagufta; Mehdi, Syed Qasim; Singh, Lalji ... Thangaraj, Kumarasamy (2017). **"Like sugar in milk": reconstructing the genetic history of the Parsi population.** Genome Biology, 18 (1), 110–110. https://doi.org/10.1186/s13059-017-1244-9.

Underhill, P.A.; Poznik, G.D.; Rootsi, S.; Järve, M.; Lin, A.A.; Wang, J.; Passarelli, B.; Kanbar, J.; Myres, N.M.; King, R.J.; Di Cristofaro, J.; Sahakyan, H.; Behar, D.M.; Kushniarevich, A.; Sarac, J.; Saric, T.; Rudan, P.; **Pathak, A.K.**; Chaubey, G.; Grugni, V. ... Villems, R. (2015). **The phylogenetic and geographic structure of Y-chromosome haplogroup R1a.** European Journal of Human Genetics, 124–31. https://doi.org/10.1038/ejhg.2014.50.

Chaubey, G.; Singh, M.; Crivellaro, F.; Tamang, R.; Nandan, A.; Singh, K.; Sharma, V.K.; **Pathak, A.K.**; Shah, A.M.; Sharma, V.; Singh, V.K.; Selvi Rani, D.; Rai, N.; Kushniarevich, A.; Ilumäe, A.M.; Karmin, M.; Phillip, A.; Verma, A.; Prank, E.; Singh, V.K. ... Thangaraj, K. (2014). **Unravelling the distinct strains of Tharu ancestry.** European Journal of Human Genetics, 22 (12), 1404–1412. https://doi.org/10.1038/ejhg.2014.36.

Rootsi, S.; Behar, D.; Järve, M.; Lin, AA.; Myres, NM.; Passarelli, B.; Poznik, D.; Tzu, S.; Sahakyan, H.; **Pathak, A**.; Metspalu, M.; Grugni, V.; Semino, O.; Metspalu, E.; Skorecki, K.; Villems, R.; Kivisild, T.; Underhill, P. (2013). **Phylogenetic applications of whole Y-chromosome sequences and the Near Eastern origin of Ashkenazi Levites.** Nature Communications, 4, Article 2928. https://doi.org/10.1038/ncomms3928.

Gazi, NN; Tamang, R; Singh, VK; Ferdous, A; **Pathak, AK**; Singh, M; Anugula, S; Veeraiah, P; Kadarkaraisamy, S; Yadav, BK; Reddy, AG; Rani, DS; Qadri, SS; Singh, L; Chaubey, G; Thangaraj, K. (2013). **Genetic Structure of Tibeto-Burman Populations of Bangladesh: Evaluating the Gene Flow along the Sides of Bay-of-Bengal.** PLOS ONE, 8, e75064.

Rai, N; Chaubey, G; Tamang, R; **Pathak, AK**; Singh, VK; Karmin, M; Singh, M; Rani, DS; Anugula, S; Yadav, BK; Singh, A; Srinivasagan, R; Yadav, A; Kahsyap, M; Narvariya, S; Reddy, AG; VanDriem, G; Underhill, PA; Villems, R; Kivisild, T ... Thangaraj, K. (2012). **The Phylogeography of Y-Chromosome Haplogroup H1a1a-M82 Reveals the Likely Indian Origin of the European Romani Populations.** PLOS ONE, 7, e48477. https://doi.org/10.1371/journal.pone.0048477.

Saxena, Shailendra K; Singh, Maneesh; **Pathak, Ajai K;** Mathur, Asha (2006). **Reply to Encephalitis outbreak finds Indian officials unprepare**d. Nature Medicine, 12, 269–270.

Saxena, Shailendra K; Singh, Maneesh; **Pathak, Ajai K**; Mathru, Asha (2005). **Attributing recent outbreak of brain fever in India as Japanese encephalitis outbreak is misleading**. BMJ, 331(7528), 1288-a.

* – *shared first authorship*

ELULOOKIRJELDUS

Nimi: Ajai K. Pathak
Sünniaeg: 1. jaanuar 1979
Perekond: elukaaslane
Kontakt: Tartu Ülikool, Genoomika instituut, Riia 23b, Tartu 51010, +372 737 5053
E-post: ajai.kumar.pathak@ut.ee

Haridustee:

2011–2019	Doktoriõpingud, Tartu Ülikool, Loodus- ja täppisteaduste valdkond, Tartu Ülikooli Molekulaar- ja rakubioloogia instituut, Evolutsioonilise bioloogia õppetool
2002–2004	Magistrikraad biotehnoloogias (*MSc*), VBS Purvanchal Ülikool, Jaunpur, India
2004	*MSc* väitekiri Raku- ja Molekulaarbioloogia Keskusest (CCMB), Hyderabad, India. Pealkiri: "Role of SULT1E1 Gene Polymorphism and Non-Random X-Chromosome Inactivation in Recurrent Early Pregnancy Loss (REPL)" juhendamisel Dr. Lakshmi Rao ja hilise Dr. Lalji Singhi juhendamisel
2003	*MSc* suvekoolitus CCMB-st, Hyderabad, India. Pealkiri: "Role of the Follistatin gene mutations in Polycystic Ovarian Syndrome (PCOS)" Dr. Lakshmi Rao ja hilise Dr. Lalji Singhi juhendamisel (Mai 2003–Juuli 2003)
1995–1998	Bakalaureusekraad (*BSc*), Purvanchal Ülikool, Jaunpur, India

Erialane tööhõive:

2018-…	Tartu Ülikool, Genoomika instituut, populatsioonigeneetika nooremteadur
2012–2017	Eesti biokeskuse külalisuurija
2006–2011	DS Genomic Solutions Pvt Ltd, India, teadlane (teadusuuringud ja IP)
2004–2006	Rakuraku ja molekulaarbioloogia keskus (CCMB), India, projektiassistent

Teaduspreemiad ja tunnustused:

2019	EMBO | EMBL Symposia reisigrant
2019	ASHG tunnustus artikli eest "The Genetic Ancestry of Modern Indus Valley Populations from Northwest India" (Trainee Paper Spotlight), August 2019
2013	DoRa T8 lühike õppevisiit
2011	Euroopa Sotsiaalfondi doktoriõppe ja rahvusvahelistumise programmi "DoRa" stipendium

Lisainfo:

2018 Valitud ja edukalt läbitud Elixir popgeni praktiline kursus "Rahvastiku genoomika: taust ja tööriistad" 21. – 27. Aprillil 2018 CNR geneetika ja biofüüsika instituudis, Napoli, Itaalia

2013 Otto Warburgi rahvusvaheline suvekool ja teadussümpoosion järgmise põlvkonna järjestuse määramise ja selle mõju kohta geneetikale, 19–26. august. 2013 Max Plancki molekulaargeneetika instituudis, Berliin, Saksamaa

2007 PG diplom patendiseadustes, NALSAR Õigusülikool, Hyderabad, India

2006 Algkursus intellektuaalomandi õiguse alal, Ülemaailmne Intellektuaal-omandi Organisatsioon (WIPO)

Loometöö:

2019 Korraldas 20. juunil Tartus joogapäeva koos India Soome saatkonna ja Eestiga, et tutvustada Ühinenud Rahvaste Organisatsiooni rahvusvahelise joogapäeva ajal joogat kui fitnessi ja tervislikke tavasid. (Tartu, Indian Minsitry of Culture & Indian Embassy of Finland & Estonia).

2018 Populatsioonigeneetikat käsitlevad teadusalased kõnelused magistrantidele, Detpartement of Bioetchnology, Faculty of Sciences, VBS Purvanchal University, Jaunpur, India

Teadusorganisatsiooniline ja -administratiivne tegevus:

Ameerika Inimgeneetika Ühingu (ASHG) liige

India Inimgeneetika Ühingu (ISHG) eluaegne liige

Teaduspublikatsioonid:

loetletud ingliskeelse CV rubriigis *Publications*

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

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

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

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

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

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

106. **Ave Suija**. Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005, 162 p.
107. **Piret Lõhmus**. Forest lichens and their substrata in Estonia. Tartu, 2005, 162 p.
108. **Inga Lips**. Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005, 156 p.
109. **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.