Realia et naturalia

DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
267

CHANDANA BASU MALLICK

Genetics of adaptive traits and gender-specific demographic processes in South Asian populations

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
267

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
267

CHANDANA BASU MALLICK

Genetics of adaptive traits and gender-specific demographic processes in South Asian populations

UNIVERSITY OF TARTU
PRESS
1632

Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu, Estonia.

The dissertation is accepted for the commencement of the degree of Doctor of Philosophy (in Molecular Biology) on 21st November, 2014 by the Council of the Institute of Molecular and Cell Biology, Faculty of Science and Technology, University of Tartu.

Supervisors: Prof. Toomas Kivisild, Visiting Professor,
Institute of Molecular and Cell Biology,
Faculty of Science and Technology, University of Tartu, Estonia
and
Reader, Division of Biological Anthropology,
University of Cambridge, United Kingdom

Prof. Richard Villems, Professor of Archaeogenetics,
Institute of Molecular and Cell Biology,
Faculty of Science and Technology, University of Tartu, Estonia

Reviewer: Prof. Maris Laan, Professor of Human Molecular Genetics, Institute of Molecular and Cell Biology,
Faculty of Science and Technology, University of Tartu, Estonia

Opponent: Dr. Sandra Beleza, Lecturer in Genetics,
University of Leicester, Leicester, United Kingdom

Commencement: Room No. 105, 23B Riia Street, Tartu, on 16th January, 2015 at 14:00.

The publication of this dissertation is granted by the Institute of Molecular and Cell Biology, University of Tartu and by the Graduate School in Biomedicine and Biotechnology created under the auspices of European Social Fund.

ISSN 1024-6479
ISBN 978-9949-32-733-1 (print)
ISBN 978-9949-32-734-8 (pdf)

University of Tartu Press
www.tyk.ee

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The current dissertation is based on the following original articles referred to in the text by their Roman numerals:

I. **Basu Mallick C***, Iliescu FM*, Möls M, Hill S, Tamang R, Chaubey G, Goto R, Ho SYW, Romero IG, Crivellaro F, Hudjashov G, Rai N, Metspalu M, Mascie-Taylor CGN, Pitchappan R, Singh L, Marta Mirazon-Lahr M, Thangaraj K, Villems R and Kivisild T. (2013) The Light Skin Allele of *SLC24A5* in South Asians and Europeans Shares Identity by Descent. PLoS Genet 9(11): e1003912. doi:10.1371/journal.pgen.1003912 (* Equal contribution)

II. Gallego Romero I, **Basu Mallick C**, Liebert A, Crivellaro F, Chaubey G, Thomas M, Villems R, Singh L, Swallow D, Thangaraj K, Mirazon-Lahr M and Kivisild T. (2012) Herders of Indian and European cattle share their predominant allele for lactase persistence. Mol Biol and Evolution. 2012; 29:249–60.

III. Chaubey G, Metspalu M, Choi Y, Mägi R, Romero IG, Soares P, van Oven M , Behar DM, Rootsi S, Hudjashov G, **Mallick CB**, Karmin M, Nelis M, Parik J, Reddy AG, Metspalu E, van Driem G, Xue Y, Tyler-Smith C, Thangaraj K, Singh L, Remm M, Richards MB, Lahr MM, Kayser M, Villems R and Kivisild T. (2010) Population Genetic Structure in Indian Austroasiatic speakers: The Role of Landscape Barriers and Sex-specific Admixture. Mol Biol and Evolution. 2010; 28:1013–24.

My contribution to the above listed articles is as follows:

Ref I – Participated in design and field work of the study, performed most of the experiments, analyzed the data and wrote the paper with contribution of other co-authors.

Ref II – Sequenced the *LCT* enhancer region in approximately half of the samples, participated in the analyses of the data and contributed to the writing of the manuscript.

Ref III – Participated in the lab work related with genotyping of the Indian samples and contributed to the writing of the manuscript.

LIST OF ABBREVIATIONS

MI	Melanin index
LP	Lactase persistence
LNP	Lactase non-persistence
LPH	Lactase-phlorizin hydrolase
AA	Austroasiatic
SNP	Single nucleotide polymorphism
bp	base pair
KYA	thousand (kilo) years ago
mtDNA	mitochondrial DNA
UVR	Ultraviolet radiation
AIM	Ancestry informative marker
AIR	Ancestry informative region
OCA	Oculocutaneous albinism
LD	Linkage disequilibrium
HGDP	Human Genome Diversity Project
XP-EHH	Cross- population Extended Haplotype Homozygosity
GWAS	Genome-wide association studies
iHS	Integrated Haplotype Score
SE	South East
T2D	Type 2 diabetes
CAD	Coronary artery disease
ANI	Ancestral North Indian
ASI	Ancestral South Indian
PCA	Principal Component Analysis

Gene acronyms used

AMY1	Salivary amylase
SLC24A5	Solute carrier family 24 member 5
LCT	Lactase
SLC45A2	Solute carrier family 45 member 2
SLC12A1	Solute carrier family 12 member 1
G6PD	Glucose-6-phosphate dehydrogenase
ATRN	Attractin
NOS2A	Nitric oxide synthase 2
EPAS1	Endothelial PAS domain-containing protein 1
TYR	Tyrosinase
TYRP1	Tyrosinase-related protein 1
DCT	Dopachrome tautomerase
SILV	Silver gene
TPCN2	Two pore segment channel 2
CTNS	Cystinosin
MYEF2	Myelin expression factor 2
MCM6	Minichromosome maintenance complex component 6
ADAM17	ADAM Metallopeptidase domain 17

I. INTRODUCTION

The origin of humans and the migratory routes they followed to inhabit different corners of the world has been a subject of research and considerable debate for biologists, archaeologists and anthropologists. Amidst these central questions which still ponder the evolutionary biologists about the human past; one of the key themes is to understand human adaptations and the various adaptive traits we have acquired during these years. There are a few distinct physical traits that differentiate us from apes (Olson & Varki 2003; Varki & Altheide 2005), our closest relatives, but humans themselves encompass huge within species phenotypic diversity (Serre & Pääbo 2004). The diverse array of phenotypes that we currently see across the globe is a result of adaptations due to various selective pressures and/or evolutionary forces acting at different points of time in the evolutionary history of our species.

The elucidation of genetic architecture of adaptive traits, most of which lie under the common umbrella of complex traits, is challenging as they are governed by both genes as well as environment and therefore, disentangling them is a crucial step in their understanding. Secondly, the environmental influences are not always constant in time and populations have often changed their environments by migration. Furthermore, culture and technology also profoundly influence the evolution of phenotypic traits (Laland et al. 2010 and references there in). Regardless of the complexity of the factors affecting phenotypic evolution, studying the molecular nature of the adaptive traits can provide us new knowledge about human diversity and potentially pinpoint the key events in the evolutionary history that have been fundamental in shaping human fitness and the present biological variation. There are two primary sources for studying human adaptations. One source is archaeology and paleoanthropological research on material culture and fossils, the second is the variation in our genomes, including modern humans and archaic hominins (Stoneking & Krause 2011).

Population genetics forms the crux of the study of genetic basis of adaptive traits and this is what has been explored in the current dissertation because the answer of "why we are different" lies in our diverse genomes. Although many adaptive phenotypic traits have been known, this dissertation in particular, highlights two specific examples – skin pigmentation and lactase persistence; their candidate genes; the causal variants, their phylogeography, selection patterns and genetic diversity among geographically distinct populations and in particular among South Asians, who behold huge array of phenotypic variation both in terms of skin color as well as varying milk drinking histories. The second part of the dissertation focuses on studying the origin of Austroasiatic (AA) speaking populations residing in South Asia. Although there has been different hypotheses proposed by linguists, historians, archaeologists and geneticists, any common consensus about their homeland that could explain the present distribution of AAs has not been achieved yet.

2. REVIEW OF LITERATURE

2.1. Understanding human adaptations

Humans show remarkable variation, both within and among populations for a number of phenotypic traits, such as height, pigmentation, altitude adaptation, cold tolerance, resistance to infectious diseases (Relethford 2009; Frazer et al. 2009). Phenotypic diversity, as we witness today, is a culmination of genetic, physical and cultural adaptations that have occurred due to various selective pressures acting over different junctures during human evolution (Balaresque et al. 2007; Hancock et al. 2010a). Phenotypic evolution has been also influenced by culture, technology and migrations (Richerson & Boyd 2008; Laland et al. 2010; Corona et al. 2013; Jeong et al. 2014). Therefore, a fine-grained understanding of the genetic basis of phenotypic traits can help us to know about human adaptations. Additionally, it can help us to understand how different selective forces have shaped the current pattern of human diversity, and how they have influenced the different biological processes, thereby reconstructing the human evolutionary past. It will also help to elucidate the population-specific variants that might be critical in understanding the differential susceptibility to complex diseases and drug response, thus putting forward foundations for genomic medicine. Furthermore, learning about adaptations will also propel the field of personal genomics which allows access to one's personal genome, as it would enable us to understand the status of one's health better–what kind of environments or diets suits us or potential risks involved for certain diseases.

Unlike the Mendelian traits which are mainly governed by single gene, most phenotypic traits are polygenic and are influenced by many loci across the genome (Lynch & Walsh 1998; Wu 2012). What makes the study even more complex, is the fact that the genes involved in phenotypic variation have different effect sizes which contribute differently to the trait (Allen Orr 1999; Allen et al. 2010), some being even too small to reach to statistical significance to be interpreted (Rockman 2012), through genome-wide association studies (GWAS). Such level of complexity requires large datasets to understand the genetic underpinnings. Besides this, to gain a comprehensive understanding of the adaptive evolution, we need information about changes in climate, human migrations in the past, cultural changes, selective forces that have acted in different time periods. Technological advancements have allowed us to measure the environmental variables with greater precision in the recent years. Therefore, understanding of adaptive traits is a trans-disciplinary theme that involves inputs from various fields (Figure 1). In addition to this, the last decade has witnessed how ancient DNA studies have been pivotal in providing clues about human phenotypic evolution (Reich et al. 2010; Reich et al. 2011; Prüfer et al. 2014; Sikora et al. 2014; Stringer 2014; Gokhman et al. 2014), as they provide a direct window to look through the past. With the advent of next generation sequencing tools and unprecedented data in conjunction with knowledge from other fields (Figure 1), we are just beginning to understand adaptive evolution

in humans. In fact, an integrated approach is the "need of the hour" i.e. to gather information from all the biological realms the "omics" – proteomics, epigenomics, metabolomics, transcriptomics, epigentics where each of these can add vital information about the understanding of the trait (Keurentjes et al. 2008). Besides this, functional genomics involving *in vivo* and *in vitro* experiments further aid in understanding the mechanistic approach of the identified candidate adaptive variants. Collaborative research holds a great promise in advancing our knowledge about the human past and in particular human adaptations. Lactase persistence (LP) as we see in later chapters is one of the best examples where evidence from genetics, archaeology, ancient DNA studies and biochemistry has been put together to attain a coherent picture of it's evolutionary past (Leonardi 2013).

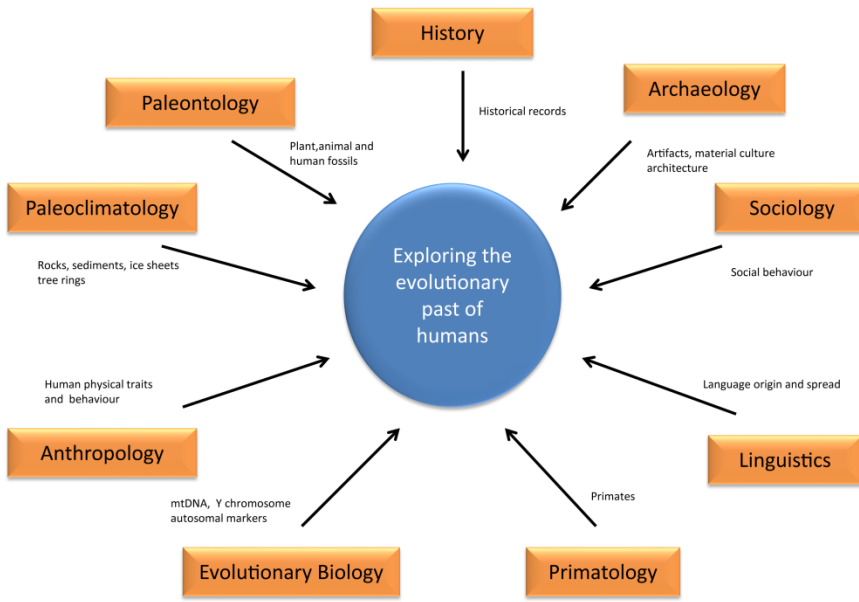

Figure 1: Exploring the human evolutionary past – a schematic model showing transdisciplinary research which involves information from various fields.

2.1.1. The crucial time epochs in human adaptation

Human adaptations have been by and large, influenced by different selective pressures acting on three different time periods in human evolutionary history since 6 million years ago (Glazko & Nei 2003; Steiper & Young 2006). Firstly during the divergence from apes and other hominids; secondly during the expansion of humans from African homeland, and thirdly Neolithic period (marked by the beginning of human settlements). Of these, latter two have been most pivotal in shaping us to what we are today (Figure 2). Comparison of the human genomes with the primates gives insight to the species-wide adaptations

over long evolutionary time scales, whereas comparison of present genomes helps us to understand the traits humans acquired in about the last 50,000 years (Harris & Meyer 2006). Therefore, a closer look through the latter time epoch (about last 50,000 years) is useful to understand about recent adaptive evolution which has contributed greatly to the within species diversity (Pritchard et al.

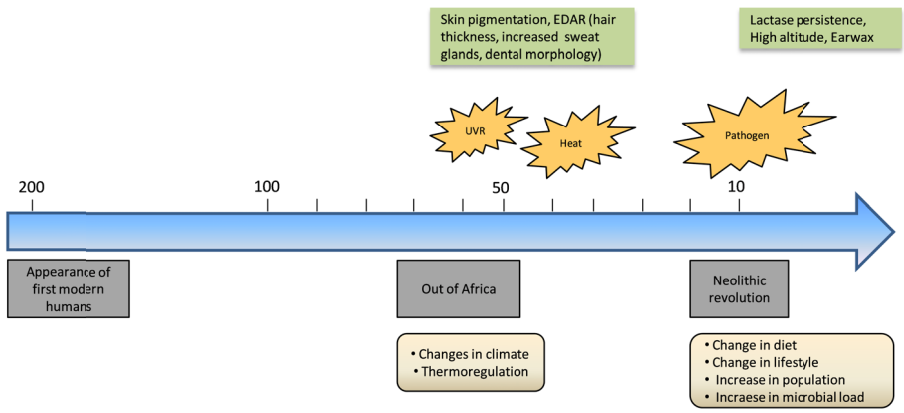

Figure 2: The blue arrow represents the schematic evolutionary timescale (in KYA); the boxes in grey represent the key events in human evolutionary history (with extended points of the specific changes that occurred during the time frame); orange stars represent the possible triggers during those periods and boxes in olive green represent the human adaptations.

Out of Africa migration (50–70 KYA): As humans dispersed out from the African homeland to conquer new land masses, they encountered many environmental challenges in the form of new habitats. In this process of colonization of different parts of the world, they faced different selective pressures for new alleles which increased their reproductive fitness and helped them to adapt in these novel environments. This favored the evolution of "adaptive alleles" and their spread across the populations. Also, in terms of climate it was not a steady one. It involved multiple oscillations of warmer and colder climates (Carto et al. 2009). The global sea level was also much lower than present times. Retraction of ice sheets provided access to the land bridges which paved the way to uninhabited lands and human expansion to wider geographic ranges (Stewart & Stringer 2012). Phenotypes pertaining to adaptation in cold environments, high altitude and skin pigmentation are most likely the examples of adaptation during this period (Scheinfeldt & Tishkoff 2013).

Neolithic transition (last 10–12 KYA): Apart from dealing with changes in the environment and climatic conditions during their expansion, humans also encountered changes in their subsistence patterns (Tresset & Vigne 2011); one

of the most revolutionary event being the transition from hunter-gatherers to a more sedentary life as agriculturists. This is marked as one of the most important developments in prehistory of mankind and termed as "Neolithic transition". This period brought forward use of polished stone tools, pottery, development of economies and appearance of human settlements (Kuijt & Goring-Morris 2002; Pinhasi et al. 2005; Renfrew 2006; Dubouloz 2008). Also, this period was characterized by dramatic shifts in diet (increase in consumption of starch and carbohydrates) (Perry et al. 2007; Luca et al. 2010; Hancock et al. 2010b) and metabolic function related adaptations influenced by domestication of plants and animals (Vigne 2011; Larson & Burger 2013). Transition from nomadic to sedentary life also led to increase in population density and microbial load which resulted in various infectious diseases as the chances of zoonotic transfer of novel pathogens was higher (Wolfe et al. 2007; Comas et al. 2013). This further led to evolution of phenotypes related to genes involved in immune response (Fumagalli et al. 2009; Fumagalli & Sironi 2014). The knowledge about these genes can be useful in adopting strategies for improvement of our health. The climate was also warm and relatively more stable after the end of last glacial maximum (~ 14 KYA), that further facilitated farming practices and population growth. Culturally, this was the first step towards industrialization and advent of modern lifestyle resulting to several societal changes like trading, increased social stratification, craftsmanship (pottery and polished stone tools), acquiring of new skills and complex economies (Bocquet-Appel & Bar-Yosef 2008). All these changes, likely resulted in a myriad of selective pressures for genotypes that were better suited to the novel environments. Hawk and his colleagues (Hawks et al. 2007), proposed that the rate of adaptive evolution gained a recent acceleration during the Neolithic period, gave rise to diverse phenotypes using a demographic model. Increase in copy number of *AMY1* gene (Perry et al. 2007) and genetic variations in *LCT* gene are important dietary adaptations of this period to high starch diet such as tubers and availability of surplus milk with the advent of dairy culture (Swallow 2003). Therefore, Neolithic transition and the sub-sequent years mark an important epoch in the evolution of important adaptive traits.

It has been also observed that some of the biological adaptations in humans that have been advantageous in the past, hold no biological relevance in the present conditions and have even turned maladaptive (Balaresque et al. 2007). For example "thrifty gene" hypothesis, which proposes that genes that enable efficient absorption and storage of fat may have been an important evolutionary adaptation for foragers during food shortage and famine, in particular for child-bearing women, may be detrimental in modern populations with abundant food sources (Neel 1962). This led to increased risk of type II diabetes, obesity and hypertension (Neel 1999; Di Rienzo & Hudson 2005; Paradies et al. 2007), as a mismatch between our genomes and present-day environments. Similarly, genetic polymorphisms that aided in salt retention, selected for humans living in hot and humid climates to balance their considerable salt loss and low dietary

salt availability, led to increased risk of high blood pressure as humans moved to temperate zones (Young et al. 2005; Balaresque et al. 2007).

2.1.2. Approaches for studying human adaptation

One approach to study human adaptations is the "candidate gene" approach where one has a *priori* information or a hypothesis, about a certain gene and its involvement with a particular trait. Second involves, use of GWAS to scan the entire genome in order to identify the putatively associated loci responsible for the particular phenotype (Figure 3). In other words, it is a blind screening of the genome to identify what the hidden text has to tell us. Till date, GWAS have been successful in identifying > 6000 variants associated with > 500 quantitative traits and disease phenotypes (Robinson et al. 2014). In principle, genome-wide scans do provide a broader picture or a starting point for understanding human adaptations. However, to interpret this catalog of information and understand its relevance, a multidisciplinary approach and in depth follow-up study involving functional and gene expression studies and subsequent validation using replication studies in independent cohorts is needed.

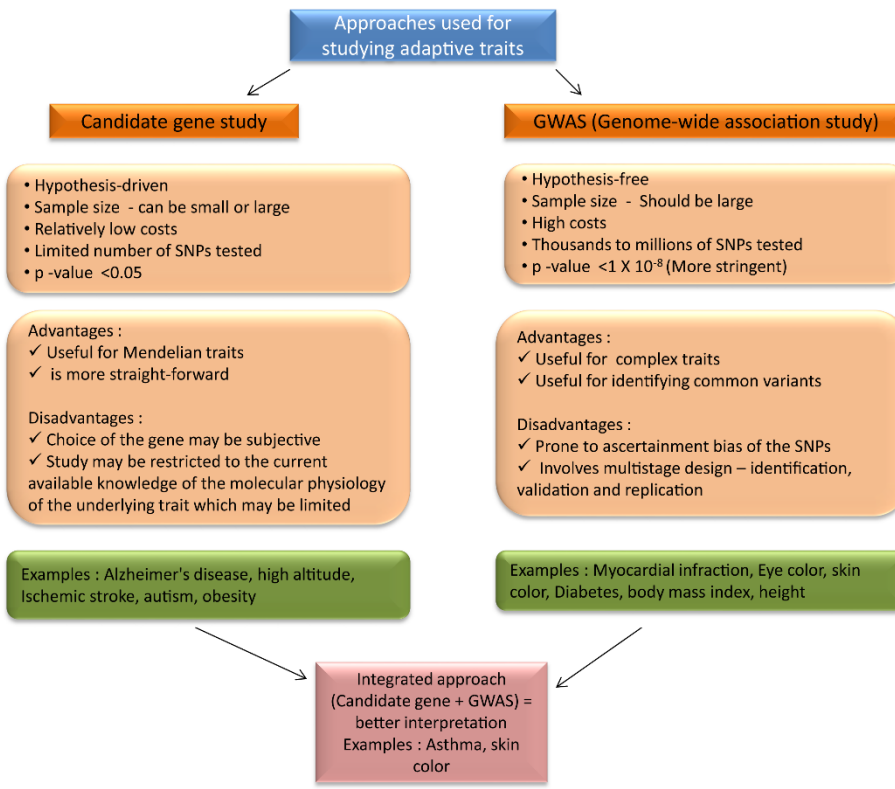

Figure 3: A comparative study of the two complementary approaches used in studying adaptive traits (Candidate gene study *vs* Genome-wide association study).

Third approach is a kind of "reverse approach" which scans the entire genome for signatures of selection of human adaptation, made feasible through SNP-based array studies for identification of adaptive loci, also referred as "reverse ecology of adaptation" (Radwan & Babik 2012). This helps in identifying genomic regions pinpointing the causative allele or the SNP being affected by natural selection. Through this, we can identify SNPs associated with adaptations to different environments, diet or disease prevalence, and how the different selective pressures have shaped the current spectrum of phenotypic variation. This approach has been also crucial in identifying recent local adaptations, as environments and selection pressures do vary from population to population and comparison of allele frequencies across the genomes provide a better genomic landscape to detect patterns of human adaptation; for example high altitude adaptation in Andeans, Tibetans and Ethiopians (Bigham et al. 2009; Bigham et al. 2010; Yi et al. 2010). Additionally, another approach to identify the signals of local adaptation is to compare the allele frequencies across different populations and correlate it with environmental variables after correction for demography (Novembre & Di Rienzo 2009; Raj et al. 2013; Frichot et al. 2013; Günther & Harttgen 2013). This allows identifying the specific environmental variable that might have acted as strong selective force. For example, diversity across the HLA class I genes suggested the selective pressure of pathogens on immunity genes (Prugnolle et al. 2005). Some studies have resulted in spectacular success stories where we have been able to identify the genetic polymorphisms, know its biological relevance, and understand the selective forces shaping them and also how it drives to reproductive fitness (Kamberov et al. 2013). But some identified adaptive loci still need a thorough investigation (Grossman et al. 2013; Forni et al. 2013). Nevertheless, next generation sequencing has and is continuing to revolutionize the study of human adaptations by producing a catalog of SNPs at the first hand, which need to be functionally annotated to make them biologically interpretable and gain a deeper understanding of phenotypic traits (Scheinfeldt & Tishkoff 2013). Whole genome sequencing allows us to study the genomes with much more precision than before, thus helping us to have a broader picture of the linked loci, and therefore holds a great promise for studying the adaptive traits in future. Furthermore, the availability of genome sequences of non human primates and archaic humans have provided important insights to the genetic basis of adaptations (Sequencing & Analysis Consortium 2005; Reich et al. 2010; Green et al. 2010; Meyer et al. 2012; Prüfer et al. 2014). In particular, some studies have proposed that gene flow from archaic humans has been fundamental in human adaptation *viz* high altitude genes (Huerta-Sánchez et al. 2014); immunity related genes *HLA* and *STAT2* (Abi-Rached et al. 2011; Mendez et al. 2012; Mendez et al. 2013) and skin pigmentation genes *BNC2* (Vernot & Akey 2014; Sankararaman et al. 2014).

2.1.3. The role of natural selection

The evolutionary paradigm of adaptation *via* natural selection was put forward by Charles Darwin in as early as 1859 but still, we do not have complete knowledge about all the potential drivers of selection or the processes that govern it. Nevertheless, the studies till date have emphasized on the central role of natural selection in shaping human adaptation which allowed humans to thrive through the extreme climate and diverse range of environments (Fu & Akey 2013 and references there in). Genomic tools and the abundance of sequence data have revolutionized the study of human population genetics and given us a great opportunity to uncover the potential signs of selection through the evolutionary history by studying the patterns of DNA polymorphisms (Nielsen 2005). The genetic variants that can be detected in the genomes of contemporary populations are crucial, in particular, for the understanding of the genetic architecture of adaptive traits. Why there was a need of a particular adaptive trait and what has been the course of its evolution in human history? Which selective pressures have shaped its current phenotypic variation? The scale of data which is available along with new computational and statistical tools that have emerged in the last decade have aided in identifying the genomic regions under non-neutral evolution i.e. selection and the comparative analysis of diverse populations helps us to understand why and how an adaptive trait emerged in a population and how it correlates with their reproductive fitness. A study involving a statistical framework testing the environmental factors such as climate, diet or pathogen load for their role in local adaptation signatures, have revealed that the selective pressure exerted by pathogens has been the dominating driver (Fumagalli & Sironi 2014). Among the 103 genes tested, most common were genes associated with autoimmune diseases such as celiac disease, type 1 diabetes and multiple sclerosis (Fumagalli et al. 2011).

Natural selection acts in three modes (Figure 4). Positive selection favors the existence of a particular allele that enhances fitness and leads ultimately to fixation, thereby decreasing genetic variation in the locus making it stand out among other genomic loci in terms of unusually high haplotype homozygosity or reduced number of segregating sites (Sabeti et al. 2006), (Akey 2009). Purifying (or negative) selection also reduces genetic variation as it eliminates deleterious mutations to maintain the integrity of functional sequences and can be recognized for example by low ratios of non-synonymous to synonymous mutations (Barreiro & Quintana-Murci 2009), (Asthana et al. 2007). Balancing selection, in contrast, increases genetic variation in a locus by favoring heterozygotes over homozygotes as it is of selective advantage for an individual to carry multiple alleles rather than one particular allele for example in immunity genes (Paterson 1998; Andrés et al. 2009; Piel et al. 2010). Each type of selection thus leaves its distinct pattern of DNA polymorphisms in our genome which can thereby be detected.

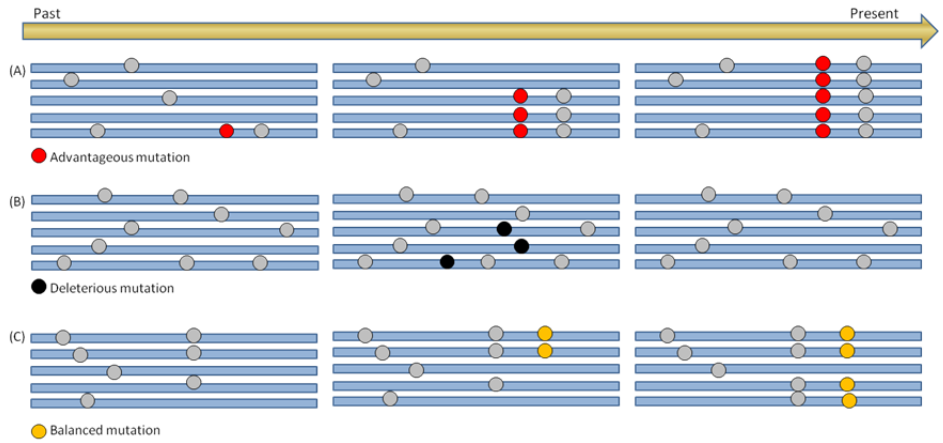

Figure 4: Figure illustrating the three modes of natural selection (A) Positive selection (B) Negative selection (C) Balancing selection.

However, if alternate episodes of selection have acted in the evolutionary history on the same genomic region, detecting them can be tedious. Population-specific signature of selection has been an important hallmark of major adaptive traits like skin pigmentation, lactase persistence, adaptation to high altitude etc. A large number of selected alleles also contribute to, or are causative of certain diseases. Therefore, presence of these signatures or footprints in the genome help to unravel functionally important genes related to complex traits or diseases (Akey et al. 2002; Ronald & Akey 2005; Sabeti et al. 2006) and provide important insights into events in the evolutionary history that have shaped the human variation (Sabeti et al. 2002; Tishkoff & Verrelli 2003). In other words, the study of natural selection marks the first steps to understand the "tell-tale" of the adaptive traits and etiology of certain diseases and complex traits.

Most importantly, looking for evidence of natural selection involves search for statistical departure from neutral expectations of molecular evolution (Tajima 1989) which states that majority of the polymorphisms have no effect on fitness or survival. These "signatures" point to subtle signs in the genome based on certain parameters; frequency of polymorphism or site frequency spectrum, genetic differentiation measures (Fst), difference between synonymous and non-synonymous rates of substitution, extent of linkage disequilibrium (LD) (reviewed in Harris & Meyer 2006). For example, a positively selected trait will exhibit increased homozygosity, reduction in haplotype diversity (Voight et al. 2006; Tang et al. 2007; Sabeti et al. 2007; Enard et al. 2014) high levels of population differentiation (Myles et al. 2008; A. Barreiro et al. 2008; Bhatia et al. 2013) or skewed allele frequency spectra (Fu & Akey 2013 and references therein) as characteristics of positive selection. Tests of selection can be based on empirical one locus vs whole genome comparisons, or

based on theoretical calculations or model-based thresholds of significance, involving computer simulations (Nielsen 2005; Kreitman 2000). As these selection tests examine the departure from neutrality through statistical approaches, they are also referred as "tests of neutrality" (Simonsen et al. 1995). It is possible that different tests of neutrality show evidence of positive selection in the same region, which then becomes cogent evidence. It can be also possible that selection varies over time, with different selective regimes during different time periods (Harris & Meyer 2006).

Nonetheless, one of the major challenges, is the fact that most of the selection tests detect recent events of positive selection because the time required either to generate or erase signals based on extended haplotypes is typically within the last 10,000 years, thus leaving behind the older signals unnoticed (Sabeti et al. 2006). On the other hand, the evidence of selection in one test and failure to detect it by other tests does not exclude the possibility of genomic selection (Sabeti et al. 2006). Another promising approach introduced by Grossman and his colleagues (Grossman et al. 2010) is a complex metric called Composite of Multiple signals (CMS) which combines different neutrality tests to a unique score which is used to detect genetic patterns of adaptation. This approach was used recently in a number of genome-wide studies to identify the putative regions of positive selection across the genomes (Grossman et al. 2013; Karlsson et al. 2014). One of the limitations in assessment of natural selection is that certain demographic scenarios tend to produce similar signatures in the genome and thus disentangling them is trivial, although attempts through new statistical methods have been made to distinguish the cofounding effects of demography (Li et al. 2011; Li et al. 2012). Another important challenge, in case of selective sweeps, where the frequency of the advantageous allele rises to fixation, is to disentangle whether it has arisen due to a single *de novo* mutation or from standing variation (Teshima & Przeworski 2006; Barrett & Schluter 2008; Peter et al. 2012). A recent study on ancient DNA has provided a method of estimating selection coefficients using a simulation based approach where selection acting on standing variation can be accommodated (Wilde et al. 2014). Till date, a large number of genomic regions have been identified as targets of positive selection using genome-wide scans (Grossman et al. 2013) or candidate gene approach (Wang et al. 2007). But only few of these regions have been shown to harbor potential functional variants or linked to an adaptive phenotype (Sabeti et al. 2007; Kamberov et al. 2013). Therefore, a comprehensive investigation of the genomic loci listed as positive selected loci, is required to gain a better understanding of human adaptation. In addition, this will help to differentiate the true adaptive loci and exclude false positives from the genome-wide scans.

Signatures of local adaptations have been also identified through whole genome sequences of specific populations (Zhou et al. 2013; Veeramah & Hammer 2014). Whole genome sequences do not suffer from ascertainment bias and also provide a broader picture of the genomic region that has undergone adaptation by capturing the entire spectrum of variation thus providing

complete characterization of site-specific spectrum. High-coverage genomes of hunter-gatherer populations (Khoe-San, Hadza, Pygmies, Sandawe) have been targeted for identifying unique signatures of selection (Granka et al. 2012; Lachance et al. 2012) as they would have evolved likely under the influence of unique environments, diets and pathogen pressures. In another study, whole genome sequencing has led to better understanding of the mechanisms underlying hypoxia tolerance among human populations inhabiting at high altitude in Andean mountains (Zhou et al. 2013).

Convergent evolution in human adaptation

It is interesting to note that some of the best studied examples of human genetic adaptation (skin pigmentation, lactase persistence, adaptation to high altitude, resistance to malaria) show existence of multiple mutations contributing to similar phenotype. This suggests that these genetic variants have followed different evolutionary trajectories to adapt to the same selective pressures posed by the environment pertaining to change in subsistence or climate or pathogen. As these are known to vary across populations and even within populations, population genetics plays a crucial role in identifying them. Independent G6PD alleles have been observed to confer resistance against *Plasmodium vivax* or *falciparum* in African, Mediterranean and Southeast Asians (Saunders 2004; Tishkoff et al. 2001; Louicharoen et al. 2009). Similarly, different alleles have been known for lactase persistence in Europeans, within Africans (East, South) and Middle East populations (Enattah et al. 2008) (Tishkoff et al. 2006), (Itan et al. 2010) and light skin for Europeans (*SLC24A5*, *MATP*, *TYR*) and East Asians (*ADAM17*, *ATRN*) (Norton et al. 2007). Another set of genetic studies point to the fact that there are distinct genes underlying adaptation to high altitude among Andeans, Ethiopians (*PRKAA1*, *NOS2A*) and Tibetans (*EPAS1*) (Bigham et al. 2013; Foll et al. 2014).

2.2. Skin pigmentation

2.2.1. Understanding the biology of the skin

Skin is the largest organ of the integumentary system and forms the first line of defense against external environment. It accounts for 15% of the body weight and plays a pivotal role in immunity, thermoregulation, homeostasis, and production of vitamin D (Lehmuskallio et al. 2002). It consists of three layers, the dermis, epidermis and the subcutaneous tissue (Figure 5). The wide array of skin color found among humans is indeed unique and interesting. The primary biological pigment that determines the color of our skin, eyes and hair is melanin. It is a mixture of polymers, mainly derived from tyrosine and is synthesized by specialized cells called melanocytes. There are two types of melanin, reddish yellow phaeomelanin and brownish black eumelanin and the ratio of these two determines the skin color (Thody et al. 1991). Within the melanocytes, melanin resides in specialized lysosome-like organelles called

melanosomes. Although the number of melanocytes might differ with the body site (Whiteman et al. 1999), the total number of melanocytes is fairly the same in all humans. Hence, what causes variation in the skin color is the difference in size, packaging and type of melanosome (Sturm et al. 1998; Young & Sheehan 2001; Robins 2005). Studies of ultrastructure of melanosomes have also revealed distinct differences between the major ethnic groups (Toda et al. 1972; Rosdahl & Szabo 1976). Individuals with lighter skin have higher pheomelanin to eumelanin ratio and their melanosomes are smaller and less densely packed (Thong et al. 2003; Lamason et al. 2005). The production of melanin involves a number of genes acting in a successive way. A four stage melanosomal maturation has been proposed starting from pre-melanosomes that bud from the endoplasmic reticulum to keratinocytes, ultimately derived from tyrosine through a number of intermediate steps (Ito & Wakamatsu 2003). Typically the skin and hair color depends on the amount of melanin in keratinocytes of upper epidermal layer, whereas those of eyes on melanosomes in retinal pigment epithelium (RPE) cells of the iris.

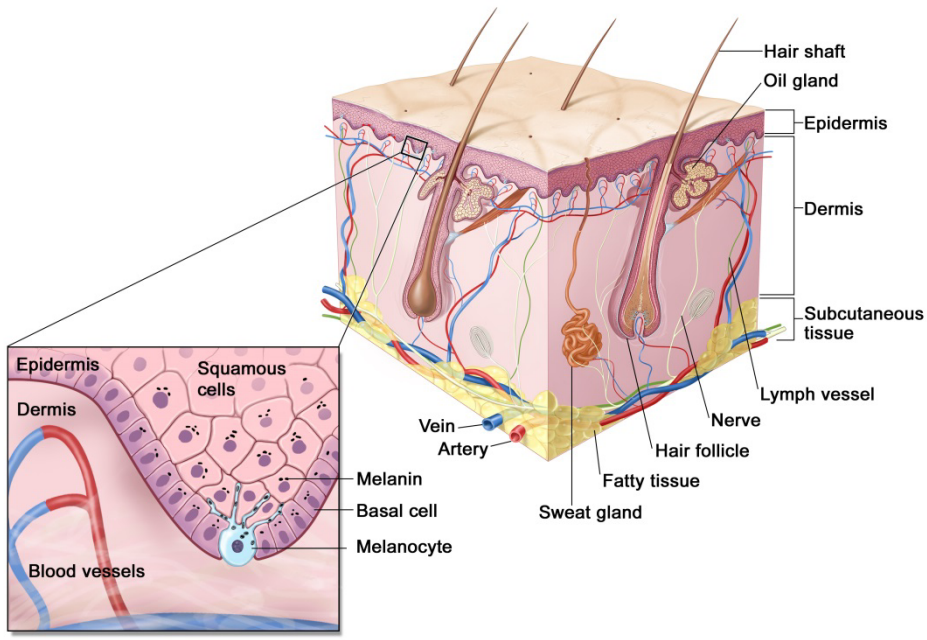

Figure 5: The basic composition of the human skin
(Image source: http://www.uchospitals.edu).

Skin pigmentation is one of the most distinct and visible phenotypic traits in humans. Studies on various genetic markers (Ryman et al. 1983; Lewontin 1995; Jorde 2000; Tishkoff & Kidd 2004) and craniometric traits (Relethford 1994), inferred that a major part of the total human variation exists within continental groups rather than among them. However, in terms of global skin

pigmentation variation, we find an opposite scenario i.e. 88% of the total variation of skin color exists among continental populations, rather than within populations (Relethford 2002).

The variety in skin color tones across the world was evident to many historians, geographers and anthropologists, but the first account of this huge variation was given by one of the Italian geographers, Renato Biasutti, who created a global skin color map (Biasutti 1967) (Figure 8). It was based on von Luschan scale, collected in the late 19th century and did not involve comprehensive sampling from all the populations. Nevertheless, it gave the first broader picture of the geographical gradient of skin color distribution. Later, a much more comprehensive compilation was made based on skin reflectance measurements (Jablonski & Chaplin 2000) clarified that the main determinant of the significant correlation of skin pigmentation and latitude is UVR (Ultraviolet radiation) intensity. Skin pigmentation, as we discuss in subsequent chapters, is mostly about baseline pigmentation also referred as constitutive pigmentation and thus should not be misinterpreted with tanning (Rubegni et al. 1999). Tanning, the process of developing darker pigmentation of the skin in response to UVR, is a form of facultative pigmentation, is temporary and subject to seasonal variation. It is in fact a process of acclimatization rather than adaptation. Study of model organisms, pigmentation disorders, and genome-wide studies have largely helped us in understanding the genetics of skin pigmentation over the last decade (Lamason et al. 2005; Miller et al. 2007; Stokowski et al. 2007; Candille et al. 2012). Till date, 378 loci, including 171 cloned genes have been catalogued in the Color Genes database (http://www.espcr.org/micemut/).

2.2.2. Quantification of skin color: From simple tiles to complex spectrophotometers

One of the optical properties of melanin is the ability to absorb, scatter and reflect light of different wavelengths (Kollias et al. 1991; Ortonne 2002). This property forms the basis of most of the non-invasive methods used for quantifying skin color. Earlier, the methods adopted for measurement of skin color were crude and subjective, and based on visual assessment. These methods also involved greater risk of inter-observer variability. Since the dawn of reflectance spectroscopy in 1950s (Weiner 1951), the accuracy and objectivity of quantitative assessment of skin pigmentation has improved significantly and the instruments available thereafter are more precise, portable and user-friendly. Besides anthropological research, where pigmentation is one of the most important and distinct human phenotypic trait, these instruments are also used by dermatologists in the diagnosis of pigmentation disorders, pigmented lesions and various skin diseases and in dermato-cosmetic research. Skin color is predominantly determined by two major pigments; hemoglobin, found in the micro vascular network of the dermis, which is responsible for the redness of

the skin, and melanin, located in the epidermis which is responsible for darkness of the skin and is photoprotective. This chapter provides an overview of the various instruments that have been widely used for quantifying the variation of skin color and how they have evolved from simple tiles to complex spectrophotometers.

Von Luschan's chromatic scale: This scale was proposed in late 1800s by Austrian anthropologist and ethnologist, Felix von Luschan. It was originally called as "Hautfarbertafel" meaning skin color board or tablet. It consists of a set of 36 opaque colored glass tiles (Figure 6) which were used to compare the unexposed area of the skin of the subject (von Luschan 1927). They undoubtedly formed the first available tool for systematically measuring skin color and was used by the anthropologists worldwide to collect skin color data (Swiatoniowski et al. 2013 and references there in). However, they were largely abandoned later (Weiner 1951; Garn et al. 1956) due to problems related to reproducibility, glare, surface imperfections, variable lighting conditions (Harrison 1957).

Figure 6: The chromatic color scale given by von Luschan (1854–1924) based on 36 glass tiles (Image credit http://www.jonathanhagos.com/files/gimgs/).

Fitzpatrick Scale: This classification system for skin pigmentation types was proposed by Thomas Fitzpatrick, a dermatologist at Harvard Medical School in 1975 (Fitzpatrick 1988; Pathak 2004). According to this classification scale, skin color has been categorized into six skin types based on sun exposure pertaining to irritation, sun sensitivity, tanning response and hyperpigmentation (Table 1). This scale is still used by some of the dermatatologists and anthropologists (Nessvi et al. 2011; Sääf et al. 2011).

Table 1: Description of the six Fitzpatrick skin types (Fitzpatrick 1988).

Skin type	Fitzpatrick skin type	Characteristics	Common nationalities
I	Very fair	freckles are common, always burns but never tans	Celtic
II	Fair	Usually burns but seldom tans.	Scandinavian
III	Fair to medium brown	sometimes mildly burns but tans gradually	Caucasian
IV	Medium to light brown	rarely burns and always tans	Mediterranean, Hispanic and some Asian
V	light to dark brown	rarely sun sensitive and always tans	Pakistani and Indian
VI	Dark brown to black	never burns but tans very easily	African

Taylor Hyperpigmentation scale: It was developed by a team of researchers at "The skin of Color Centre", St Luke's Roosevelt, New York. It consists of 10 uniquely laminated plastic cards with ten different shades of skin color (A–J), each having further 10 gradations thus providing 100 possible combinations to match to the skin of the subject (Taylor et al. 2006). The major advantage lies in the fact that it is easy to use, quick and requires minimal training. It also covers all the six Fitzpatrick skin types. As the skin hues are restricted to 15 cards it may not be able to capture the large variation in skin color and therefore it is not much used. However, it can be useful to assess hyperpigmentation.

Reflectance spectroscopy: It covers a range of instruments that are based on reflectance spectroscopy, a modern method which replaced the previous visual skin color matching methods. The advantage offered by these instruments was that unlike color charts and tiles, they provided a method to detect the continuous variation of skin color and hence were more appropriate for genotype-phenotype studies. Furthermore, they were easy to use, portable and involved higher precision in skin color measurements compared to above described methods. The first reflectance spectrophotometer of its kind was E.E.L instrument (Evans Electroselenium Co., Halstead, Essex,UK) used by Weiner (Weiner 1951; Byard 1981) and later by Jablonski (Jablonski & Chaplin 2000). Most importantly, reflectance spectrophotometers readings were reproducible and were not affected by inter-observer error (Lees & Byard 1978).

Tristimulus colorimetry: The principle behind this instrument lies in the fact that it is analogous to the way our eye perceives the color. It uses white light to illuminate the skin and then measures the intensity of the reflected light through three filters of particular wavelength or photodiode arrays (Clarys et al. 2000). The results are expressed using Commission International d'Eclairage (CIE) lab system, a standard method for colorimetry, a 3 digit output in which colors are expressed in terms of their lightness (L*) , amount of green or red (a*) and amount of yellow or blue (b*) (Westerhof 1995; Takiwaki et al. 2002) thus

recording the color in a three dimensional space. Most widely used Tristimulus colorimeter instruments are tristimulus colorimeters, Chromameter (Minolta Chroma Meter CR-200, Osaka, Japan) and ColorWalk. Chromameter illuminates the skin using a pulsed xenon arc lamp and then measures the intensity of reflected light through filters of 450, 560 and 600 nm wavelengths. ColorWalk on the other hand uses photodiode arrays. One of the major advantages is the availability of reliable and reproducible data (Kerckhove et al. 2003).

Narrowband spectrometers: They consist of light-emitting diodes to illuminate the skin surface and the intensity of reflected light is recorded using a photodetector. Most widely used narrowband spectrophotometers are, DermaSpectrometer (Cortex Technology, Hadsund, Denmark), Erythrema/ Melanin Meter (DiaStron, DiaStron Ltd., Hampshire, UK) and Mexameter M X16 (Courage Khazaka). The highlight of these instruments is that they can distinguish between two main pigments of the skin *viz.* melanin and hemoglobin using their distinct spectral curves (Diffey et al. 1984). DermaSpectrometer uses two light-emitting diodes, green (568nm) and red (655nm) to illuminate the skin and the amount of the reflected light is measured using a photodetector (Park 2002). Similarly, Mexameter uses 16 light-emitting diodes which emit light at three wavelengths–green (568 nm), red (660 nm) and infrared (880 nm). Results are expressed in terms of "melanin index (MI)" as quantitative measurement of melanin and "erythrema index (E)" as quantitative measurement of redness of skin. Thus, the melanin content of the skin as defined by Parra (Parra 2007).

$$M \text{ (melanin index)} = \log 10 (1/ \% \text{ red reflectance})$$

The main idea is that hemoglobin shows a large peak at green wavelength with little absorption in the red wavelength and hence reflectance of a narrowband light at the red spectrum would give the likely measure of the melanin. An average of 3 readings is recommended for the instrument. Nevertheless, these instruments offer reliability and reproducibility of the measures as well as ease of use and portability.

Diffuse reflectance spectroscopy: It appears to be most versatile and non-invasive method of measuring skin pigmentation (Stamatas et al. 2004). It is based on the principle that the reflectance of the skin for the entire visible spectrum is measured at a high optical resolution. The availability of data from full spectrum thus allows even better differentiation of the two chromophores; melanin and hemoglobin, showing absorbance at 620-720 nm and 560-580 nm respectively. So, besides the advantage of acquisition of data from the whole spectrum, these new generation spectrophotometers also provide a uniform platform thus allowing to calculate all type of pigmentation measures; i.e. CIELab system, M and E index etc. The common one is Stellarnet EPP2000 reflectance spectrometer (Stellarnet, Tampa, Florida).

To summarize, we have now instruments that can objectively measure the continuous variation in human skin color which is suitable for genetic studies taking into account the quantitative nature of the trait (Shriver et al. 2003), (Jacobs et al. 2013). In Ref I, we have used one of these above mentioned instruments, a narrowband spectrophotometer; DermaSpetrometer (Cortex Technology, Hadsund, Denmark) to record the skin pigmentation readings among South Asians.

2.2.3. Evolutionary history of skin color

Though comparative genomics and association studies have revealed hundreds of pigmentation genes (http://www.espcr.org/micemut/), our understanding about which pigmentation genes and exactly at what phase of our evolutionary history, went through adaptive changes is still not complete. Skin is one of the distinct features of our phenotype. It's a reflection of a person's age, health and ancestry (Jablonski 2004). Therefore, understanding the evolutionary history of skin color *i.e.* the journey that led to the diverse array of skin tones is indeed an interesting challenge for population geneticists. Even Charles Darwin himself was enthralled at the wide variation of skin color he observed during his voyage (Darwin 1871; Darwin 2004). He mentioned this as:-

> "*Of all the differences between the races of man, the color of the skin is the most conspicuous and the best marked…*"

It has been hypothesized that the skin of the first hominids was probably lighter than present sub-Saharan Africans, although covered with dark body hair similar to chimpanzees (Post 1975; Jablonski & Chaplin 2010). As a part of human evolution, one of the key processes, that allowed our ancestors to survive well in tropical environments was loss of body hair and increase in the number of eccrine sweat glands, especially around face and for protection of brain, which is heat sensitive (Cabanac & Caputa 1979; Jablonski et al. 2006). Thus, the loss of body hair left the skin exposed to different environmental conditions including strong UVR and dark skin developed thereafter, as an adaptation that marked the beginning of evolution of skin pigmentation (Jablonski & Chaplin 2000). Since then, the human skin color has evolved multiple times independently and has been largely influenced by natural selection (Norton et al. 2007; Jablonski & Chaplin 2010). The genetic evidence emerging from the previous studies, coherently points to a complex picture with multiple episodes of selection involving various genes, at different time periods in the evolutionary history with some genes being selected only in some populations (Figure 7) (McEvoy et al. 2006; Quillen & Shriver 2011). As the genes leading to light skin were not the same for Europeans and East Asians, it was apparent that skin pigmentation underwent convergent adaptation with different genes leading to the similar phenotype (Norton et al. 2007).

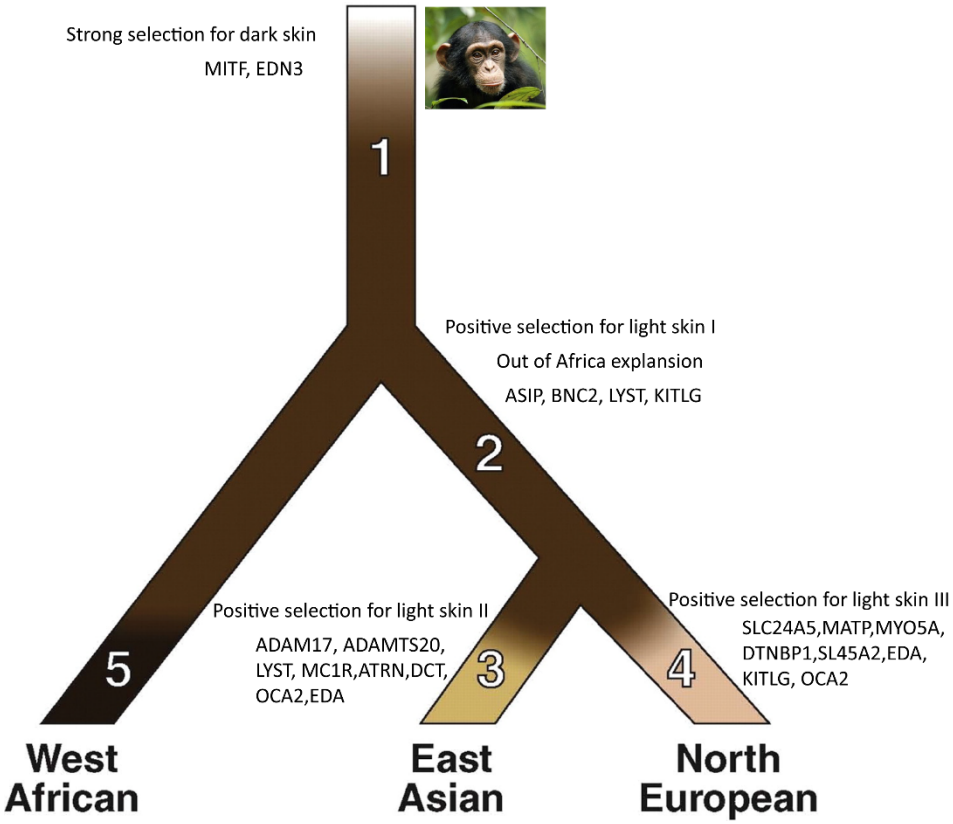

Figure 7: A schematic framework for evolutionary model of skin pigmentation (adapted from (McEvoy et al.2006; Quillen and Shriver 2011) studied among three HapMap populations (Africans, East Asians and Europeans). Pigmentation trends are indicated by shading of the branches. Branch numbers show the different episodes of selection and the genes mentioned at each branch are the ones which show evidence of positive selection among 77 candidate pigmentation genes studied.

It is apparent that the major driving force that has led to the geographical stratification of skin color is adaptation to different UVR conditions; however this cannot completely explain the entire skin color diversity. The role played by "other evolutionary forces", in particular demography, which might have also contributed to the biological spectrum of skin color, has not been yet fully elucidated but is already reflected in few studies (Norton et al. 2007; Beleza et al. 2013; Sturm & Duffy 2013). In this chapter, I will briefly review the current state of knowledge about the different hypotheses put forward to explain the present global skin color variation. We delineate hypotheses which have strong genetic evidences and others are theoretical and do not have thorough explanations yet.

Different hypotheses put forward to explain overall variation

The latitudinal cline

The exploration towards the cause of skin color variation has indeed a long history. Hippocrates in the fourth-fifth centuries BC has speculated that variations in human traits including skin color was related to environment (Isaac 2006). Thereafter, Aristotle further accentuated the fact by proposing "climatic theory" according to which dark skin was attributed to intense sunshine and heat. In the 18th century, John Mitchell and Samuel Stanhope, two naturalists, observed that skin pigmentation follows a clinal pattern with latitude (Figure 8) (Mitchell & Collinson 1744; Smith 1965). Smith was confident that it was some uniform factor in the climate (mainly related to sunshine) that caused the latitudinal cline but he wasn't able to pinpoint it exactly. This fact was attested by a number of studies thus suggesting a strong correlation between skin color, latitude and intensity of UV radiation (Roberts & Kahlon 1976; Tasa et al. 1985). Later, it was demonstrated that skin reflectance measures were in particular related to annual average erythemal dose of UVR, mostly UVB (Jablonski & Chaplin 2000; Chaplin 2004; Jablonski & Chaplin 2010). Here, skin color measurements were compared exactly with the global satellite data of UV intensity. An independent study (Relethford 2000) confirmed that this gradient was not similar in both hemispheres and there are exceptions to this general clinal pattern.

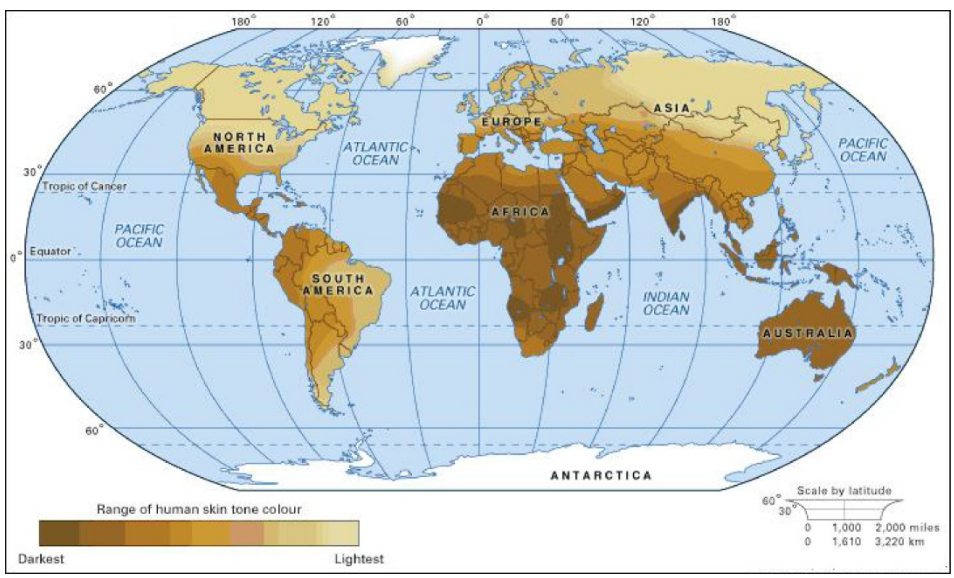

Figure 8: Skin color distribution map across the world showing variation with latitude. (Image credit: In *Encyclopædia Britannica*, Retrieved from http://www.britannica.com).

Sexual selection

The idea of sexual selection was first proposed by Darwin (Darwin 1871) who believed that it might play a substantial role in skin color variation among humans. One of the studies suggested that the preference of lighter-than-average skin for mate choice could have been one of the factors influencing the pigmentation variation in humans (Van den Berghe & Frost 1986). However, some authors argue that if this had been true, sexual dimorphism should have increased with increasing distance from equator, which is not the case (Madrigal & Kelly 2006), while some others advocate that it may have played a secondary role (Aoki 2002). A recent genetic study showed that sexual selection has a convincing role in phenotypic variation among human faces (Claes et al. 2014). However, it's role in skin color variation remains yet to be rigorously shown.

Evolution of dark skin

When our ancestors lost their body hair, their skin became vulnerable to harmful effects of UV radiation. Melanin, thereby acted as a "natural sun-screen" and evolutionary acquisition of dark skin in regions of high UVR by our ancestors can be seen as an adaptation to protect the skin against sun-burn or skin cancer and/or from degradation of folate (Branda & Eaton 1978; Jablonski 1999). Folate is pivotal for DNA synthesis, cell division and repair and also plays a key role in spermatogenesis (Cosentino et al. 1990; Ebisch et al. 2006). Therefore, deficiency in folate levels can result in anemia, complications on pregnancy, and multiple fetal abnormalities involving neural tube defects (Bower & Stanley 1989; Fleming 2001), some of which can be even lethal. Comparison of dark and light skinned individuals suggested latter to be more prone to photolysis of folate when subjected to UV rays (Branda & Eaton 1978). Hence, dark skin would have been advantageous in high UVR environment to maintain optimal folate levels. It has been argued that as skin cancer has late-onset and is not fatal (Blum 1961), so the protection of nutrient photolysis, in particular folate, could have been a stronger driver for the evolution of the dark skin. Nevertheless, some albinos living in the equatorial region do get affected by skin cancer early in their life (Okoro 1975).

Evolution of light skin: The role of vitamin D

Besides their damaging role UV rays are also beneficial, in particular UVB, for synthesis of vitamin D. Although we can find dietary sources of vitamin D, its primary source is cutaneous synthesis on exposure to sun (Holick 2003; Holick 2005; Holick & Jenkins 2005). The main precursor of vitamin D is 7-dehydrocholesterol which is present in the keratinocytes of the epidermal layer of the skin. On activation by UVB rays, provitamin D3 is formed which is further converted to biologically active form of vitamin D essential for regulating the calcium and phosphorus levels, development and mineralization of bones and regulation of cell growth (Grant & Holick 2005; Holick & Chen 2008). It is believed that as humans moved to northern latitudes where they

encountered lower levels of UVR which proved to be a disadvantage for sufficient production of vitamin D. As a consequence, individuals with lighter skin have gained advantage since this phenotype allows higher penetration of UVB rays, enhancing thereby the vitamin D synthesis (Loomis 1967). According to this original hypothesis proposed by Loomis (Loomis 1967), dark skin was selected in regions close to equator to prevent excess of vitamin D which could be toxic, whereas selection favored light skin in regions away from equator to maximize the production of vitamin D. However, Holick (Holick 2007) disproved the toxic effects of vitamin D, as production of vitamin D is a rate-limiting process. This led to the conclusion that the geographical distribution of skin color is not solely dependent on vitamin D levels. Hence, the revised version of the hypothesis, which is also the most plausible explanation till date is that, the skin color distribution is a balance between the skin to be dark enough at high UVR zones to protect against sun burn and degradation of folate and light enough at low UVR zones to allow sufficient production of vitamin D. In other words, the consensus states that skin pigmentation is a result of adaptation to different UVR conditions *via* natural selection.

It has been hypothesized that probably agriculture or settling down of modern humans resulted in a strong selective pressure for light skin evolution because of associated shift of diet from animal-rich food (rich in vitamin D) to carbohydrates (grains). This dietary shift may have led to diseases related to vitamin deficiencies and hence depigmentation of skin may have become a pivotal fitness increasing phenotype. However, direct evidence to support this view has still been scarce or lacking. However, the equivocal thought that emerged from the genetic studies regarding evolutionary history of skin is that depigmentation or acquisition of light skin was complex and multi-step process in human history (Figure 7). Genetic evidence suggests that light skin has evolved at least twice in human history (Norton et al. 2007) but dark skin probably only once (Lao et al. 2007). One involved proto-Eurasians and the other two involved Europeans and East Asians, which most likely occurred after the separation of ancestors of these two lineages (Canfield et al. 2013). However, the timing of these selection events for depigmentation, in the course of human evolution is also not clear, which might be crucial in understanding the evolutionary history of the trait. It appears that the drive for efficient vitamin D synthesis cannot be the sole factor for depigmentation.

Evidence from ancient DNA

Sequencing of the ancient genomes of West European farmer (from Stuttgart, ~ 7,000 years) (Lazaridis et al. 2014), Scandinavian stone age farmer (*Gokhem* 2, ~5000 years) (Skoglund et al. 2014) and Chalcolithic Tyrolean Iceman (~ 5,300 years) (Keller et al. 2012) have demonstrated presence of derived allele (rs1426654 SNP of *SLC24A5*, discussed in detail in Chapter 2.2.4). On the other hand, Paleolithic Siberian *MA1* (~ 24,000 years) and foragers from Luxemourg (*Loschbour*, ~ 8000 years) (Lazaridis et al. 2014), Scandinavia (*Ajvide 58*, ~ 5000 years) (Skoglund et al. 2014) and Spain (*La Brana*, about 7,000 year old)

(Olalde et al. 2014) have been accounted for presence of ancestral alleles. However, one of the hunter gatherers of the 7 analyzed (*Motala 12*) by Lazardis and his colleagues (Lazaridis et al. 2014) showed presence of a copy of derived allele, thus indicating that Mesolithic Europeans likely had both the lighter and darker variants of the allele.

Therefore, even more than a decade of research in pigmentation genetics, there is still a significant gap in knowledge concerning the evolutionary history of skin color, in particular about the exact mechanism behind skin lightening and its timing. It is also not known if the different hypotheses stated above have acted independently at different times, or the present spectrum of skin color is a cumulative effect of all or few of them in different time periods during the evolution of modern humans. Therefore, with the recent methodological improvements, synthesis of both modern and ancient DNA studies would be pivotal in enlightening us about the dynamics of skin color evolution and spread of adaptive light skin color alleles in Eurasia.

Sexual dimorphism of skin color

Most studies that have quantitatively assessed skin pigmentation in human populations have consistently revealed sexually dimorphic patterns, with mostly females being lighter than males (Tobias 1978; Relethford et al. 1983). Many evolutionary explanations have been put forward to explain this sexual dimorphism. These differences in skin color could be resultant of the differential influence of hormones estrogen and androgen on the melanocyte activity (Garn 1956); differential exposure to sun (Edwards & Duntley 1939; Harrison & Owen 1964); clothing habits (Harvey & Lord 1978) or higher needs of females of vitamin D during lactation and pregnancy (Jablonski & Chaplin 2000). Van den Berghe (Van den Berghe & Frost 1986) on the other hand, argued that sexual selection itself could be one of the primary causes of sexual dimorphism of skin color because males universally favor light skinned females as their partners. As there hasn't been any genetic evidence in support of the proposed hypotheses, the actual cause for sexual dimorphism in skin color remains elusive.

2.2.4. *SLC24A5*: A special focus

Discovery of the gene

Serendipity has influenced some of the important scientific discoveries. Keith Cheng and his colleagues, while working on cancer using zebra fish as a model organism, accidently discovered that one of the mutations in *SLC24A5* led to lighter stripes (Lamason et al. 2005), which eventually led to the identification and characterization of the human ortholog, being fundamental in evolution of light skin in humans. This gene, as the name suggests, is a member of solute carrier (SLC) family and is located on Chromosome 15q21.1 and codes for a protein NCKX5, which is a potassium-dependent sodium/calcium exchanger.

Association with skin pigmentation in humans

A non-synonymous mutation in the third exon of this gene has been known to explain the highest pigmentation differences (25–38%) between Europeans and West Africans (Lamason et al. 2005). Ginger (Ginger et al. 2008) in her study showed partial localization of NCKX5 in trans-Golgi network suggesting that this gene is directly involved in production and/or processing of melanosomes (Figure 9). Later, it was suggested that it also plays a important role in skin pigmentation variation among South Asians living in UK (Stokowski et al 2008), Orang Asli tribes of Malaysia (Ang et al. 2012) and African-European admixed populations of Cape Verde islands (Beleza et al. 2013).

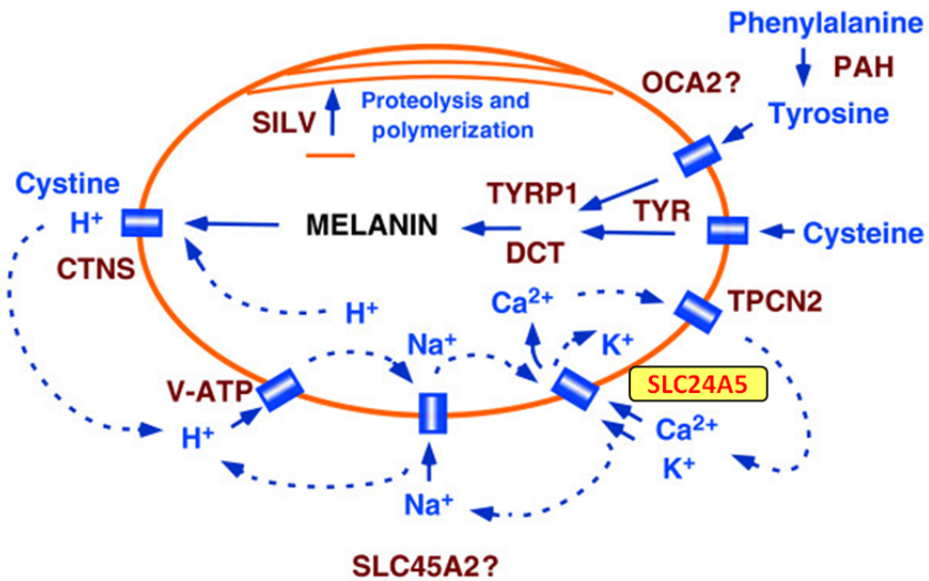

Figure 9: A melanosome showing the different genes involved in melanogenesis highlighting the gene under study (Ref I) (Adapted from Sturm et al. 2012). *SLC24A5* is one of the genes of the sodium dependent potassium/calcium exchanger family. It is involved in ion transport via V-ATP complex as shown here which is critical for the funtioning and/or processing of melanosomes.

Signature for positive selection

Comparison of sequences from four populations (Yourbas, Chinese, Japanese and Europeans) (Lamason et al. 2005; Altshuler et al. 2005) revealed that there is a sharp decrease in heterozygosity in Europeans in 150kb region of chromosome 15 which included *SLC24A5*, *MYEF2*, *CTNX2* and part of *SLC12A1,* which pointed to the evidence of strong positive selection acting around this region. The phylogeography of the rs1426654 SNP (Norton et al. 2007), also

shows that the derived allele is nearly fixed among Europeans extending further to some populations of Middle East and Pakistan. A genome-wide scan of over 3 million polymorphisms from the International HapMap Project Phase 2 highlighted *SLC24A5* as one of the strongest candidates of positively selected genes among Europeans (Lamason et al. 2005; Voight et al. 2006; Sabeti et al. 2007). Since then, many studies have confirmed the genetic evidence of positive selection (Tang et al. 2007; Pickrell et al. 2009; Coop et al. 2009; Grossman et al. 2010; Li et al. 2013) among Europeans and in Middle Eastern and Pakistani populations (Pickrell et al. 2009; Coop et al. 2009).

Age estimates for evolution of light skin

With the functional studies confirming the role of *SLC24A5* in reduction in melanogenesis *via* trafficking of pre melanosomes (Ginger et al. 2008; Tsetskhladze et al. 2012), genotype-phenotype studies (Stokowski et al. 2007; Durso et al. 2014) and evidence of positive selection, *SLC24A5* has been ascribed as one of the most important genes in the evolution of light skin. Hence, age estimates of the mutation (rs1426654) contributing to light skin would be vital for understanding the dynamics of skin evolution, in particular when and where did it originate? Previous studies providing age estimates for the rs1426654 SNP has been mostly restricted to estimation of the onset of selective sweep rather than coalescence age of the mutation. A recent study by Beleza and her coworkers (Beleza et al. 2012), estimated that selective sweep at *SLC24A5* occurred around 11-19 KYA. Nevertheless, the exact origin and timing of the SNP (rs1426654) conferring light skin remains still to be understood. Below, I will discuss the fact that *SLC24A5* is a multifaceted gene and its implications beyond pigmentation variation. Besides playing an important role in evolution of light skin, it has also been highlighted as important AIM (ancestry informative marker) in some other studies and causative of oculocutaneous albinism (OCA).

Other suggested roles of the gene

Role of *SLC24A5* as AIM

SLC24A5 plays a pivotal role in the inference of the ancestry and phenotypic characteristics of the perpetrator in forensics, crucial for the criminal investigators (Kayser & de Knijff 2011; Butler 2014). This has led to the emergence of the field of Forensic DNA Phenotyping (FDP), i.e. ability to predict the externally visible characteristics (EVC) of an individual using the material found under the crime scene with the help of DNA markers (Liu et al. 2013). The rs1426654 SNP of *SLC24A5* ranks as the second most differentiated SNP worldwide among 3011 ancestry informative markers (AIMs), and has been used extensively along with other markers to determine the ethnic biogeographic ancestry of an individual (Phillips et al. 2007; Giardina et al. 2008; Bouakaze et al. 2009). The practical application is based on the fact that the functional SNP (rs1426654) shows marked contrasting frequency differences

between European and African populations (Soejima & Koda 2007). It has been also used in conjunction with Y chromosomal SNPs to distinguish different ethnic groups among US sub-populations (Sims & Ballantyne 2008). Additionally, Giardana and his colleagues (Giardina et al. 2008) suggested that use of rs1426654 together with flanking SNPs rs2555364 and rs16960620 can increase the effectiveness of the prediction and thus proposed the idea of using ancestry informative regions (AIR) over AIMs in ancestry prediction. rs1426654 has also been an integral part of multiplex SNP assays used for prediction of ancestry (Jia et al. 2014; Phillips et al. 2014) and phenotype prediction (Valenzuela et al. 2010; Spichenok et al. 2011; Pneuman et al. 2012; Hart et al. 2013; Dembinski & Picard 2014) or both (Gettings et al. 2014).

SLC24A5 and its association with OCA

Oculocutaneous albinism (OCA) is a autosomal recessive disorder characterized by reduction or complete loss of melanin in the skin, hair and eyes and is often accompanied with symptoms like strabismus, moderate-to-severe visual impairment, photophobia, and nystagmus (Tomita & Suzuki 2004; Grønskov et al. 2007). The classification of OCA subtypes is mainly based on the causative gene rather than on clinical symptoms. Till date, 7 subtypes have been known involving genes *TYR* (OCA1), *OCA2* (OCA2), *TYRP1* (OCA3) and *SLC45A2* (OCA4) and an unidentified gene in the region 4q24 (OCA5) (reviewed in Montoliu et al. 2014). Although albinism affects individuals worldwide, it is interesting to note that the prevalence of different subtypes varies among populations of different ethnic backgrounds (Gargiulo et al. 2011). The role of *SLC24A5* in hypopigmentation has been suggested earlier in model organisms resulting in albinotic features and milder reduction of melanosome size and pigmentation in mouse B16 melanocytes (Vogel et al. 2008) and dysmorphic melanosomes in zebrafish (Lamason et al. 2005), although analyses of genetic variation in *SLC24A5* did not reveal any pathological mutations in OCA patients (Grønskov et al. 2009; Sengupta et al. 2010). However, the possibility of *SLC24A5* as putative candidate gene for OCA in humans has been explored only recently. A 4 bp insertion in *SLC24A5* gene (c.569_570insATTA) was observed in a cutaneous hypopigmented Indian individual (Mondal et al. 2012). Exome sequencing of *SLC24A5* in an OCA affected family led Wei and his colleagues to identify *SLC24A5* as a candidate gene for non syndromic OCA and classified it as OCA6 (Wei et al. 2013). The individual with OCA phenotype was characterized having two deleterious mutations c591G>A in exon 6 and c.1361insT in exon 9 of *SLC24A5*. The clinical symptoms of the OCA6 patients further tested by Fanny (Fanny et al. 2013), and his colleagues in seven OCA6 patients showed that though being rare, it is found in people of diverse ethnic origins and is not restricted to only Chinese population. On examining the morphology of the melanosomes in epidermal melanocytes, it was found that the patients have less mature melanosomes that apparently leads to decrease in the amount of melanin, thus resulting in hypopigmentation.

2.3. Lactase persistence: Ability to digest milk

Milk is considered as an everyday drink and a perfect example of balanced diet consisting of fat, protein, carbohydrate, vitamins, calcium, minerals and water (Lee et al. 1979). A glass of milk is a common element of our breakfast tables, although many people are not aware that humans do vary in their ability to digest milk. Human milk contains lactose (a disaccharide) and enzyme lactase (LPH, lactase-phlorizin hydrolase) present in the small intestine, helps to break it down to monosaccharides glucose and galactose (Figure 10A), which is then readily absorbed by the gut (Swallow 2003). At birth, all of us have the ability to digest milk because of the continuous active production of lactase, which allows us to digest mother's milk. However, after weaning, when the child's diet changes from that exclusively based on milk to various independent sources of food, there is a gradual reduction in the levels of the enzyme. This inability to digest lactose post weaning is called as lactase-non persistence (Swallow 2003) or lactose intolerance or adult-type hypolactasia or primary hypolactasia in medical terms (Sahi 1994). Individuals who have lactose intolerance have health complications on consumption of fresh milk including symptoms like intestinal gas, bloating, nausea, vomiting, flatulence, severe abdominal pain and diarrhea (Simoons 1970a; Järvelä 2005). These symptoms are developed mainly due to fermentation of undigested lactose by colonic bacteria, leading to production of fatty acids and gases (Figure 10B). Most humans are lactose intolerant and therefore cannot digest lactose in adulthood. The lactose intolerant individuals can, however, digest processed milk products like cheese, butter, and other products with ease, which have lesser levels of lactose (Alm 1982).

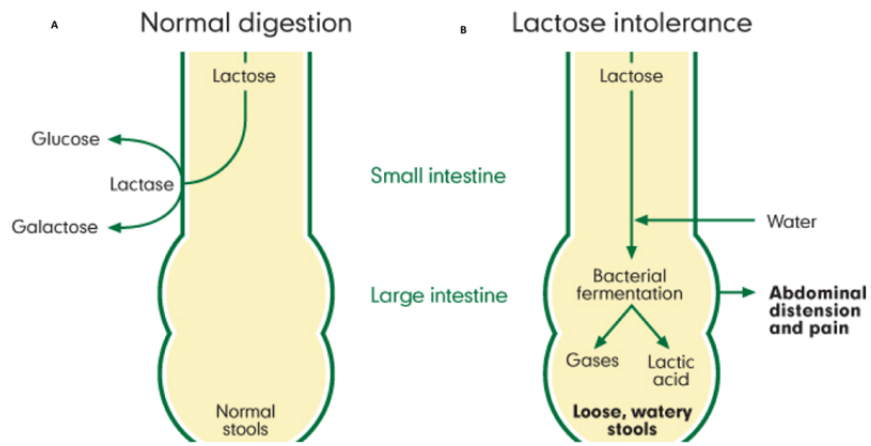

Figure 10: A schematic description of physiology of (A) lactase persistent and (B) lactase non persistent individuals.
(Image source: http://foodallergynottingham.co.uk/foodallergynottingham.co.uk).

However, some of us are "lucky" (35% of the world population) (Ingram et al. 2009) in maintaining the ability to digest lactose even in adulthood (Scrimshaw & Murray 1988), and can enjoy drinking milk throughout our life due to continued production of the enzyme (Ingram et al. 2007) and this is referred as lactase persistence (LP). Interestingly, this trait is highly variable among different populations of the world, both among and within continents (Ingram et al. 2009; Itan et al. 2010). It is high among Northern Europeans (> 90% in Swedes and Danes) and certain nomadic tribes of Africa and Saudi Arabia, about 30–70% in Middle East, South and Central Asia and rare among Native Americans and East Asians (~ 1% in Chinese). The unequivocal view brought forward from the studies done till date suggests that this trait is a consequence of dietary adaptation, brought forward when humans started domesticating animals and thus correlates well with the spread of dairying culture (Hollox et al. 2001; Myles et al. 2005). LP alleles, i.e the mutations that led to continuous production of lactase became advantageous for humans as milk consumption increased. Those individuals who carried these mutations were able to survive better as milk provided a constant source of nutrition than those who couldn't. Unlike skin pigmentation, the geographic distribution of this trait doesn't follow a distinct pattern. In Europe and South Asia it follows a Northwest to Southeast cline (Swallow 2003; Ingram et al. 2009), although in Africa the distribution is quite patchy (Mulcare et al. 2004), with high frequencies observed in pastoralist populations (Tishkoff et al. 2006; et al. 2009). It has been suggested that both demographic factors as well as natural selection have influenced the pattern of variability across the world populations (Poulter et al. 2003; Hollox et al. 2001; Gerbault et al. 2011). Studies on the genetic basis of this trait have mostly pointed at the evolution of the trait in conjunction with pastoralism suggesting gene-culture coevolution (Laland et al. 2010).

How lactose intolerance is phenotypically determined?

There are three main tests conventionally used to determine lactose intolerance phenotype. Lactose hydrogen breath test (which measures the increase in the level of hydrogen in breath due to fermentation of undigested lactose by bacteria), lactose tolerance (which measures the increase in the amount of glucose in blood) and stool acidity (that measures the levels of lactic acid in stool). Most of these common methods used to diagnose lactose intolerance are tested after ingestion of 50 gm of lactose after 8 hours of fasting. A more direct and reliable approach, although invasive, is to look for lactase activity by intestinal biopsy but it is not recommended for children (Bayless 1971; Biller & Grand 1990).

2.3.1. Genetics of lactase persistence

Though the heritable autosomal dominant nature of the trait was revealed as early as 1960s (Sahi 1973; Sahi & Launiala 1977), the exact causal genetic variant of the lactase persistence trait was not known until the last few years. The gene encoding lactase or lactase-phlorizin hydrolase, *LCT* (OMIM 223100), is located on chromosome 2 (Figure 11) and was first discovered in late 1980s (Kruse et al. 1988). Interestingly, scanning through the exons and promoter region (1kb) of the *LCT* gene among lactase persistent and non-persistent individuals failed to identify any putative variants (Mantei et al. 1988; Boll et al. 1991), to be associated with the trait. Although some haplotypes were identified, the exact causal SNP remained elusive. Nevertheless, the difference in the lactase mRNA levels between LP and LNP individuals suggested involvement of a cis-acting element which is responsible for the down regulation of the lactase gene (Wang et al. 1995; Wang et al. 1998). Finally, the starting point was made by Enattah and her colleagues when they discovered -13910T and -22018 G/A to be strongly associated with lactase persistence in a Finnish cohort (Enattah et al. 2002). These SNPs were positioned within the 13^{th} intron and 9^{th} intron of the neighboring gene *MCM6* (Figure 11). Further, *in-vitro* functional assays confirmed, that indeed the variant -13910T had a pivotal role in the up regulation of the *LCT* (Kuokkanen et al. 2003; Troelsen et al. 2003; Olds & Sibley 2003; Lewinsky et al. 2005), and -22018G/A is in close LD with the causal SNP (Bersaglieri et al. 2004; Rasinperä et al. 2004; Ranciaro et al. 2014). This finding also further bolstered the importance of regulatory mutations in human adaptation (Wray 2007; Visser et al. 2012; Fraser 2013). As the genetic association of LP with the identified SNP was 100% in the Finnish populations it appeared that the -13910 C/T single nucleotide change alone could explain majority of the lactase persistence cases. However, the follow-up studies testing the association of -13910T SNP in culturally and geographically diverse African populations and Middle East populations revealed that this theory doesn't hold true for all lactose tolerant populations (Mulcare et al. 2004; Bulhões et al. 2007; Tishkoff et al. 2006; Imtiaz et al. 2007; Enattah et al. 2008). -13910 C/T, the primary determinant of lactase persistence among Europeans was absent among most of the African populations including East African pastoralist populations (Mulcare et al. 2004). This raised questions about what explains the distribution of the trait in other lactase persistent populations. To explore this, Tishkoff (Tishkoff et al. 2006) studied the region in the vicinity of the -13910T variant in search of their potential variants amongst 470 individuals from 43 ethnic groups from Tanzania, Kenya and Sudan which allowed identification of 4 new LP alleles -14009, -14010, -13907 and -13915 (located within 100 bp of -13910) (Figure 11), to be distinct causal variants associated with lactase persistence. This led to important evidence that LP alleles evolved independently in Europeans and Africans owing to convergent evolution due to strong selective pressure (Tishkoff et al. 2006; Enattah et al. 2007; Ranciaro et al. 2014).

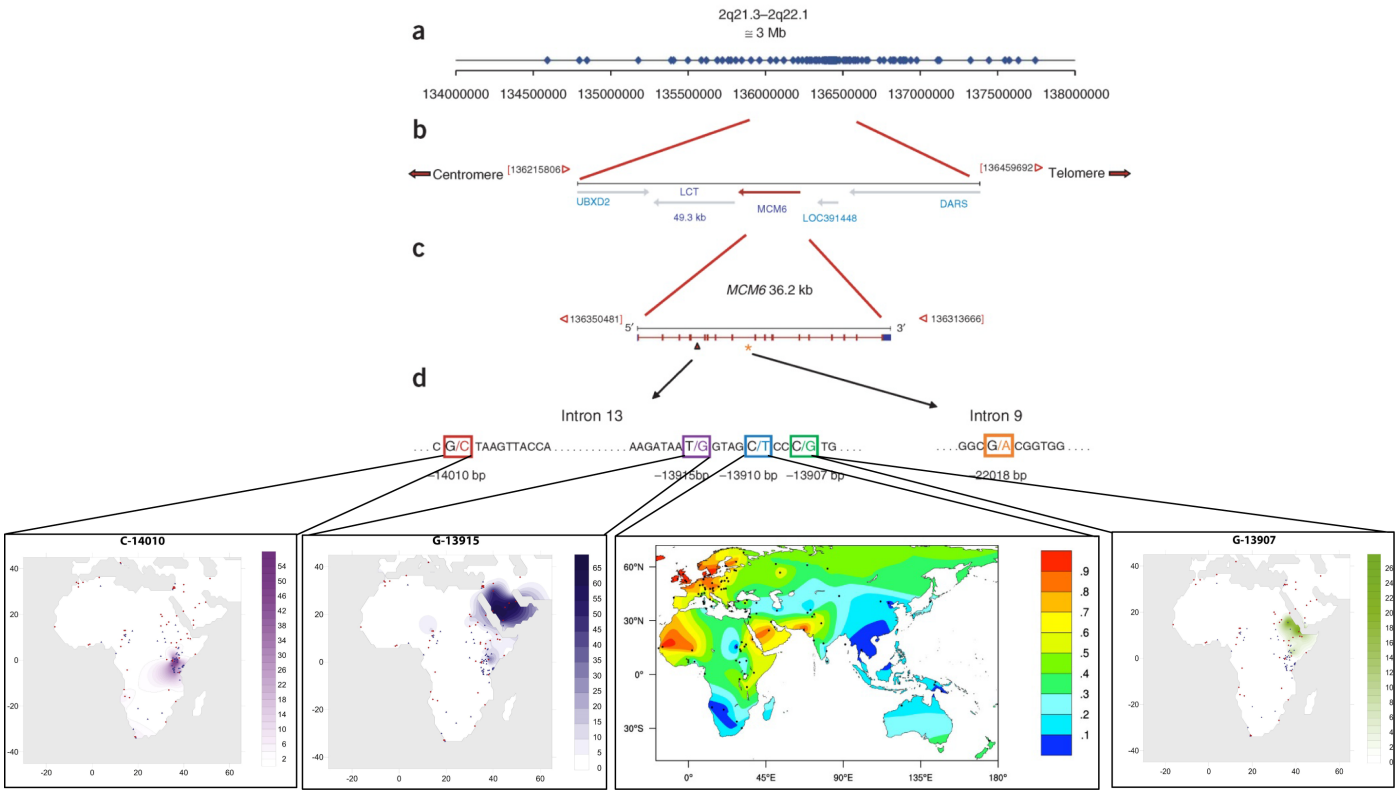

Figure 11: Chromosomal location of *LCT* gene and various genetic variants of lactase persistence (Adapted from Tishkoff et al. 2006; Itan et al. 2010; Ranciaro et al. 2014).

Compared to the newly identified variants which have more or less restricted geographic distribution in Africa and Saudi Arabia, -13910T has a wider geographic distribution extending from Europe, Central Asia, China, South Asia and some North, Central and/or South African pastoralist populations (Mozabite, Arabic Baggara, Fulani, Nama, Bulala and Hausa) (Heyer et al. 2011; Ranciaro et al. 2014; Bersaglieri et al. 2004). The presence of the variant in African populations and observed haplotype variation suggests that this allele was most likely introduced in them by gene flow from outside Africa (Tishkoff et al. 2009; Henn et al. 2012), or migration events within Africa or recent admixture (Breton et al. 2014). Another study comparing the Tibetans and Han Chinese revealed that lactase persistence was significantly higher in Tibetans (30%) than in Han Chinese (3.3%), consistent with their milk consumption habits (Simoons 1970a). When the genetics of lactase persistence was explored among Tibetans, it was found that none of the previously detected polymorphisms for the trait was present among the individuals studied (Peng et al. 2012). Nevertheless, one of the newly identified polymorphism -13838G/A present in appreciable frequencies (1.9 to 20.8%), thereby suggesting that lactase persistence could have independent origin in Tibet (Peng et al. 2012), although it's exact functional role is still to be ascertained. It was further suggested that the current global distribution of lactase persistence has been influenced by trajectories followed by different LP alleles mainly *via* natural selection, thus leading to their rapid spread and have been also subsequently reintroduced in some places through migrations and local genetic admixture (Ranciaro et al. 2014; Breton et al. 2014). It is also possible that as only few pockets of pastoralist populations have been targeted for the search of LP variants; there could be some variants which are not known yet.

2.3.2. Tracing the "milky past"

The genetic history of lactase persistence (LP) is fascinating. We know that the LP alleles enhanced the efficiency of digesting milk and thus increased the chances of better survival. Therefore, positive selection has favored the individuals having the beneficial mutation. But, even after a decade of research on genetics of lactase persistence, still the knowledge about the evolutionary history of lactase persistence is incomplete. So far, it is not known where and when did the different LP alleles originate and what has been the central selective agent? It is therefore interesting to explore the genetic underpinnings of the "milky past" i.e. how this trait evolved in different populations and how they have led to the adoption of the present diet regime followed by different populations. Some studies argue that as a constant supply of fresh milk was a prerequisite so it is highly unlikely that the trait was at appreciable frequencies before dairying. According to this model, in particular defined as the "cultural historical hypothesis" the LP alleles were rare and dairying provided a strong selective advantage to the humans who could digest fresh milk and thereby leading to its rise to higher frequencies in some populations via positive

selection (McCracken 1971; Beja-Pereira et al. 2003; Gerbault et al. 2011). On the other hand, the "reverse cause hypothesis" states that these mutations were already common in certain populations due to random genetic drift, they were the ones who adopted the practice of dairying (Nei 1985; Aoki 2001) later. Therefore, the presence of lactase persistence determined the milk adoption and consumption practices. The first hypothesis is also well supported by the subsequent ancient DNA studies which suggest that lactase persistence was absent or rare in early Neolithic farmers (Burger et al. 2007; Malmström et al. 2010). In the following chapter I review the current state of knowledge about the evolutionary history of different LP alleles.

The varying LP alleles in different populations shows that it is yet another example of convergent evolution pointing out the fact that the LP alleles arose independently and have followed parallel evolutionary paths in several geographic/ethnic groups resulting in heterogeneous distribution of the LP alleles across the world populations. The first evidence of convergent evolution appeared when Tishkoff and her colleagues identified the novel causal variants of lactase persistence in pastoralist tribes of Africa and observed that they do not share the same haplotype background as that of the genetic variant in Europeans (Tishkoff et al. 2006). Looking back to the history, convergent evolution has been a common and integral element of the human adaptations be it skin color or malarial resistance or high altitude (Norton et al. 2007; Mangano & Modiano 2014; Foll et al. 2014). This further bolsters the fact that humans have undergone similar selective pressures – in diet or adaptation to environment that have converged to the similar phenotypes across the world, thus pointing how crucial these evolutionary changes were for human adaptation.

Evolutionary history of different LP alleles

The question whether the origin of LP alleles predates dairying and pastoralism has been addressed with different methods and approaches. One of the earlier attempts to date the selective sweep around the main causative allele in Europe (-13910 C/T) based on the decay of long-range haplotypes suggested the age of 2,188 to 20,650 years (Bersaglieri et al. 2004). Another study by Coelho (Coelho et al. 2005) and his colleagues based on microsatellites associated with the T allele estimated the age of accumulated genetic variation in the range of 7,450–12,300 years. Though the first of these estimates shows an overlap with the archeological evidence of introduction of cattle to Europe about 10,750–10,250 years ago (Bollongino et al. 2006; Magee et al. 2014), it was still not clear whether LP was an adaptation to Neolithic transition or LP defining mutations were already present before the Neolithic. Biochemical analysis of fat residues found in the pottery has provided evidence for dairying practices from different Neolithic sites throughout Europe. The first evidence of processing or fermentation of milk was provided by Evershed (Evershed et al. 2008). Similarly, other evidence of use of dairy products has been evident in Neolithic sites of Romania and Hungary (7900–7450 years), Britain (6100 years) and Scotland (3000 years).

Evidence from ancient DNA

Ancient DNA studies in Neolithic populations from different parts of Europe have revealed variable frequencies (0–27%) of -13910T allele (Burger et al. 2007; Malmström et al. 2010; Lacan et al. 2011; Plantinga et al. 2012). Ancient DNA from eight Neolithic skeletons of Central, Northeast and Southeast Europe (7,000–8,000 years) showed the absence of the LP allele thus suggesting that this mutation was not wide spread among the early Neolithic humans (Burger 2007). Similar evidence was obtained from a settlement site from Southern France (dated 7000 years), showing absence of the derived allele (Lacan et al. 2011). However, 10 skeletons analyzed from a Middle Neolithic site (dated 4800-4200 years) in Gotland suggested a T allele frequency of 5% (Malmström et al. 2010). Additionally, ancient DNA analysis of 46 individuals of South-west Europe from Late Neolithic-Chalcolithic period (dated 4500-5000 years) excavated from Basque Country, revealed unusually high lactase persistence of 27% (Plantinga et al. 2012). Therefore, it was concluded that lactase persistence was most likely nearly absent among the Early and Middle Neolithic and evolved later when pastoralism provided a strong selective force for its rapid spread across populations.

Besides -13910T other mutations in the lactase enhancer region have been detected. It has been suggested that -13915G which is common in the Arabian Peninsula and in African pastoralist populations with Arabian ancestry (Beja from Sudan, Arabic Baggara from Cameroon) originated about 4,095 (± 2045 years) ago possibly as a result of domestication of camel (6000 years) in Arabian Peninsula (Cook 1978; Enattah et al. 2007; Imtiaz et al. 2007) and that this allele was introduced to East Africa within the last 1,400 years with spread of Islam (Ingram et al. 2007).

-14010C, the variant prevalent in East and South African pastoralist populations (Tishkoff et al. 2006; Coelho et al. 2009; Torniainen et al. 2009) was dated to 2,700-6,800 years (95% CI 1,200–23,000 years) which brackets the timing of pastoralist migrations to eastern Africa.

-13907G has been found among Cushitic-speaking populations from Sudan, Kenya and Ethiopia (Tishkoff et al. 2006; Ingram et al. 2007; Ingram et al. 2009; Enattah et al. 2008). As it has been mainly restricted to these populations, the most likely origin would be Ethiopia.

Although we do have a comprehensive list of LP alleles, they do not account for all the phenotypic variation observed worldwide (Ranciaro et al. 2014). There are populations, for example Hadza (hunter-gatherer), who are highly lactase persistent (47%), determined phenotypically by LTT (Tishkoff et al. 2006) and yet do not carry any of the known LP variants nor have any history of animal domestication. On the other hand, there are individuals who carry LP alleles, yet are LNP (determined phenotypically). The latter situation of mismatch between genotype and phenotype can be readily explained by symptoms being related to intestinal infections other than lactose intolerance *per se* (Ranciaro et al. 2014), but the former still needs a genetic explanation. One possibility is that they may have other genetic variants which are not known yet or they contain microflora in their gut that help in digestion of lactose. Hence,

these populations could have their signatures of lactase persistence not in their genomes but in their microbiomes. A recent study comparing their microbiomes with other populations revealed higher diversity (Schnorr et al. 2014) among them, although their exact role in lactose absorption is still not clear.

Evidence of positive selection and the "primary" selective force

LCT has been among the top notch candidates of strong positive selection being revealed time and again by various genome-wide and candidate gene studies (Sabeti et al. 2006; Voight et al. 2006; Akey 2009; Grossman et al. 2010). The first evidence of positive selection among Europeans was given by Bersaglieri and his colleagues (Bersaglieri et al. 2004), who observed an unusually long haplotype extending over >1 Mb among Europeans. In the following years, another study marked the presence of extensive LD around -14010C with haplotype homozygosity extending up to > 2 Mb (Sabeti 2007). The selection coefficients for these LP variants estimated in the above studies were typically high (0.014–0.19 and 0.01–0.15 respectively), compared to other alleles that have undergone positive selection for example G6PD (0.02–0.05) (Tishkoff et al. 2001) and therefore *LCT* represents one of the strongest signatures of positive selection.

Though evidence from archaeology and allele frequencies procured from the ancient DNA samples favor the cultural-historical hypothesis, but the reasons why LP trait provided a strong selective advantage resulting in its rapid spread in the last 10,000 years remains elusive (Gerbault 2014). It also appears that LP alleles evolved multiple times in the human evolutionary history in different populations (Figure 11). This means that it would have conferred a "special advantage". Though an equivocal thought has been achieved that trait of lactase persistence has evolved due to the selective pressure induced by the advent of dairying culture, it is still not understood how this trait conferred to biological fitness in the pre-historic times.

Milk has high nutritive value being important source of carbohydrate, fat, protein, calcium, minerals and essential vitamins (Lee et al. 1979) and can be a constant source of food. Individuals who could digest this "super food" may have had a fitness advantage over those who could not digest milk (Simoons 1970a). Second, it could be a ready source of food during times of crop failure or famine and thus was succumbed to episodic selective pressures during these times. Third, it could be a source of pathogen-free drink as it contains high water content (87–88%) and could be useful in arid climates as the only source of clean water (Cook & Al-Torki 1975; Cook 1978). Although the latter characteristic can be true for Africa, it may be less relevant for temperate climate of Europe (Gerbault et al. 2011). Fourth, it could be useful for people in Northern latitudes living in low UV areas as it would facilitate absorption of lactose and reduce the risk of rickets and osteomalacia in adults (Flatz & Rotthauwe 1973) more readily explained as "calcium assimilation hypothesis". However, a recent study (Sverrisdóttir et al. 2014) argued that this may be insufficient to explain the spread of lactase persistence in Europe.

To summarize, the adaptive trait of lactase persistence highlights an example of recent evolution shaped both by genes and culture and also provides a model of how information from different scientific disciplines including genetics, archaeology, ancient DNA etc have been useful in coherent understanding of its evolutionary past (Leonardi 2013).

2.3.3. Milking history of India

India is one of the world's largest producers and consumers of milk (Simoons et al. 1970b). We do not know exactly when domestication of animals and use of milk started in South Asia due to lack of archaeological evidence. Nevertheless, existing evidence suggests that Indian subcontinent has been the original homeland of two important milking animals *viz* domesticated Zebu cattle (*Bos indicus*) and water buffalo (*Bubalus bubalis*) (Naik 1978; Patel & Meadow 1998; Kierstein et al. 2004; Kumar S. et al. 2007). Seven Neolithic centres have been proposed in South Asia (Fuller 2006) of which Baluchistan region (Pakistan) and Indus valley region are thought to be primary centres of origin of Zebu cattle, while South India being a secondary centre (Allchin & Allchin 1974). Excavations around Harrapa and Mohenjo-daro suggest the presence of zebu cattle, a distinct species of South Asian origin (characterized by a hump) in Indus valley region about 5000–8000 years ago (Meadow 1993). This was further supported by similar evidence from genetic studies (Chen et al. 2010) based on mtDNA. On the other hand, it has been also suggested that Baluchistan region was the homeland of South Asian auroch *Bos primigenius namadicus*, the purported ancestor of Zebu cattle (Grigson 1991), (Bradley & Magee 2006) and some of them survived to give rise to the latter. Most of the milk consumed in India comes from water buffalo (51%), cow and goat and three major milk producing states are Uttar Pradesh, Rajasthan, Andhra Pradesh (http://www.nddb.org/). About 45% milk is consumed raw (Singh & Pundir 2000) while the rest is used in processed forms such as in preparation of traditional sweets, cheese, paneer (Indian cheese), yoghurt, butter, ghee (Indian clarified butter) and hence milk forms a substantial part of Indian diet since historic times. The ancient epics Mahabharata and Bhagavad Gita have also mentioned butter as the favorite food of Lord Krishna and thus practice of domestication of cattle in prehistoric times (Hawley 1979; Hawley 1981). Additionally, milk forms an important source of nutrition (vitamin B12) for vegetarians (31%) living in India (Naik et al. 2013).

Pastoralist tribes of India

Pastoralism is defined as a subsistence of living where people are mainly dependent on livestock. Most of the pastoralist groups are nomadic as they move from place to place in search for better pastures for their reared animals. Unlike pastoralists of Africa and Middle East who are organized in tribal groups and live separately, Indian pastoralists form an integral part of the Indian

society in the form of endogamous social groups and are well integrated into the caste system. Different types of livestock include cow, buffalo, goat, sheep, camel, donkey and yak. Cow is also considered as one of the sacred animals and hence domesticated more for milk production rather than meat. These pastoralists inhabit wide range of niches ranging from arid parts of Thar dessert, alpine and sub-alpine zones of Himalayas, Deccan plateau, Western Ghats. According to a survey, more than 200 tribes comprising of 6% of the population are engaged in pastoralism (Sharma et al. 2003). Indian pastoralists can be divided into mainly two groups – those which practice horizontal movement in the dry region and those who involve vertical movement i.e. from lowland to highland and vice versa. For example in Himalayas, as soon as the snow melts they move to highland in summer in search of more abundant pastures and lush green fields with tender grass. Starting from September they start descending to the foothills. Therefore, transhumance (cyclic migratory pattern of herds between fixed points to exploit seasonal availability of pastures) forms an integral part of the pastoral communities of the Himalayas (Bhasin 2011).

Indian subcontinent harbors a large number of pastorlist groups. Some of the distinct ones include Golla and Kuruma of Andhra Pradesh, Raikas of Rajasthan, Bharwads of Gujarat, Gujjars of Northwest India, Rors of Haryana, Toda of Nilgiri mountains, Dhangras of Maharashtra, Bhotias, Sherpas, Monpas, Changpas of Trans-Himalayan region. For most of them pastoralism represents an inherited profession from their ancestors (like Raikas from Rajasthan as traditional camel pastoralists) and hence holds a great responsibility while for some it is the only source of subsistence in the existing region.

Lactase persistence in India

Although having an immense history of milking and pastoralism, we do not know much about the lactase persistence trait in the Indian subcontinent, i.e. whether the Indians have the unique ability to digest milk in adulthood. If yes, then it would be interesting to know which LP alleles they carry. When and how they acquired it? Do they share some of the LP alleles that are found in Eurasia and Africa or they have some novel LP alleles which are not known yet, possibly having independent origin of the trait in the subcontinent, explaining majority of the lactase persistence in India? Do the LP alleles correlate with the dairying culture and is the trait at higher frequencies among pastoralists?

Phenotypic studies, very limited though, suggest a moderately high frequency (highest frequency recorded as 0.73) of lactase persistence over the subcontinent and demonstrate a North-South cline (Desai et al. 1970; Swaminathan et al. 1970; Tandon et al. 1981; Gupta et al. 2007; Reddy & Pershad 1972), but there has been significant lack in studies related to the genetics of the trait. Therefore, studying the dairy diaspora among South Asians is an important cornerstone for understanding genetic architecture of the trait and its evolutionary history. In Ref II, we have made the first attempt to decipher the genetics of lactase persistent trait in India, accounted different LP alleles found, studied their phylogeography and selection patterns.

2.4. South Asia

South Asia refers to the southern geographic region of Asia. It is surrounded by Himalayan fringes in the North, Indian Ocean in the southern periphery and Iranian plateau in the West. The geographical boundaries of South Asia include India, Pakistan, Bangladesh, Nepal, Bhutan, Sri Lanka and the Maldives. It has been suggested that South Asia, served as incubator pot for human founding populations that migrated out of Africa (Macaulay et al. 2005; Thangaraj et al. 2006; Oppenheimer 2012), and therefore forms the first major "staging post" in the dispersal of modern humans and their settlement across the corners of the world. This proposal has been augmented by mtDNA analyses (Kivisild et al. 2003; Basu et al. 2003; Endicott et al. 2003; Metspalu et al. 2004; Palanichamy et al. 2004; Thangaraj et al. 2006; Chandrasekar et al. 2009), suggesting presence of deep autochthonous lineages in South Asia and archeological evidences (Mellars 2006, Mellars et al. 2013). On the other hand, some studies argue that there was no incubation, but an early dispersal route from Northern Africa to Eurasia through Levant during the Upper Paleolithic era about 45 KYA (Lahr & Foley 1994; Prugnolle et al. 2005). Migrations, admixture, demography, social structure have certainly influenced the genetic structure of South Asian populations, resulting in the formation of a complex demographic profile, where restricted gene flow and several evolutionary forces have generated a high level of genetic differentiation among various caste and tribal populations and also made them distinct from other continental populations (Kivisild et al. 2003; Metspalu et al. 2004; Chaubey et al. 2007; Reich et al. 2009; Xing et al. 2010). Anthropology, linguistics, classical genetic markers and modern molecular studies have contributed largely to the knowledge of the existing Indian genetic heritage (Bamshad et al. 2003; Kivisild et al. 2003; James et al. 2005; Chaubey et al. 2007; Reich et al. 2009; Xing et al. 2010). With a home for more than one billion of people where ethnic groups representing different cultures, languages, modes of subsistence and phenotypes have co-existed, South Asia serves as a perfect model to study genetic variations shaping the human diversity. This chapter focuses on the features or elements that contribute to the genetic diversity of South Asia and describes how it serves as an important paradigm to understand more about human evolutionary past and is crucial for studying varying phenotypic traits.

2.4.1. Factors shaping the genetic diversity of the Indian subcontinent

Language

Most of the languages spoken in South Asia belong to four major language families–Indo-European, Dravidian, Austroasiatic and Tibeto-Burman (Figure 12). About 73% of the populations speaking Indo-European languages mostly reside in the North and Northwest parts of the subcontinent. Dravidian speakers are mostly confined to four southern states of India, although with some

exceptions like Kurukh, Gondi and Malto branches spread to other parts of South Asia. The Tibeto-Burman speakers are spread along the Trans Himalayan range and in North-East part of India with exception of Balti population from Pakistan. Austroasiatic (AA) speakers are mainly confined to the central and eastern parts of the country in the form of small pockets interspersed with populations, speaking other languages.

Figure 12: Map of India showing the distribution of major language families.

Geography

South Asia has a variety of landscapes ranging from arid lands, deserts, forests, fertile plains, plateaus, swamps, coastal shores, deep bays, high mountains and lowlands (Field et al. 2007). These environmental challenges can be also crucial for shaping the genetic diversity, thereby conferring some unique adaptations among the populations.

Society

South Asian society is well structured consisting of several castes, tribes and religious groups. In totality, it comprises of 4635 well defined population groups including 532 tribal among which 72 are primitive tribes, out of which 36 are still hunter-gatherers (Tamang & Thangaraj 2012). Social structuring in South Asia is further fabricated by strict practice of endogamy that allows inter-ethnic separation of different groups although sharing similar socio-cultural practices and rituals, thus limiting gene flow outside caste barriers.

Caste system

Indian caste system is a unique entity in itself and forms an integral part of the Indian society. South Asia has been also known as the "land of castes". Caste is a kind of a tag name that gives a social and cultural identity to a person by birth, mostly related to people from Hindu religion. Strict endogamy where people living in the same region or of same social status maintain their unique socio-cultural identity and do not exchange genes, forms the main backbone of Indian caste system. This has given rise to a multitude of endogamous pockets thus creating higher genetic differentiation even across small geographical distances. The traditional caste system is mainly based on the doctrine of "Chaturvarna" where "Chatur" means four and "varna" refers to classes or categories. The main basis of classification was by the actions people performed in the society and so that each of them can help the society function well as a unit. Brahmins were mainly involved in teaching and performing religious rites of the society. Kshatriyas were warriors or defendants of the society. Vaishya were the artisans, traders and cultivators. Shudra mainly was involved in labor activities.
Endogamy is also practiced in other communities worldwide, but South Asian endogamy is exceptional of its kind as it harbors multiple layers. Each caste is further divided into sub-castes which are divided into smaller units called gotras (exogamous clans). In principle, gotra consist of individuals who are supposed to be descended from the common sage-ancestor. In other words, gotra speaks about the unbroken lineage and the link to the common male ancestor. Gotra also is fundamental for performing any traditional rituals, for example celebration of birth of child, making offerings in the temple, during funeral ceremonies etc. It is a common practice to inquire about one's gotra during marriage, as marriage within the same gotra (*Sagotri* marriage) is not allowed and considered as a taboo. According to the laws of "Manu" there should be a minimum separation of 7 generations between the bride and the groom (Naegele 2008), thus strictly prohibiting consanguineous marriages. However, some of the castes of South India do practice consanguineous marriage (Bittles et al. 1993; Krishnamoorthy & Audinarayana 2001). In short, group endogamy and clan exogamy constitute the marriage boundaries. Castes in particular have more rigid boundaries in comparison to tribe.

Thus, this division of castes into sub-castes then to gotras, makes the social system of South Asia quite diverse and complex (Watkins et al. 2008). However, it is of great interest for geneticists to study how this structured

classification of the social system and inter-ethnic separation between various endogamous groups has contributed to the genetic diversity of the subcontinent.

Impact of migrations

The prehistory of India accounts a number of invasions and rulers from different parts of the world ranging from Near East to Central Steppes to Southeast Asia (Bamshad et al. 2001; Diamond & Bellwood 2003; Witzel 2005). The historical records covering up to three millennia provide a catalog of these invasions. These invasions offered the possibility of a multitude of cultural contacts during these years. However, the impact of these invasions on the genetic structuring of the populations and associated gene flow is still an open debate. Indo-Aryan invasion theory, proposed initially by German orientalist Max Müller, states that around 3,500 years ago, a dramatic migration took place which also brought with it Indo-European language and caste system, and therefore forms a substantial part of their heritage. However, genetic evidence based on matrilineal, patrilineal and autosomes in the studies carried out till date have not been able to provide any clear-cut evidence of major episode of gene flow (Kivisild et al. 2003; Chaubey et al. 2007; Metspalu et al. 2004; Reich et al. 2009; Metspalu et al. 2009; Xing et al. 2010). On the other hand, studies reflecting minor gene flow have been evident. One clear example of migration which is supported by the genetic evidence is that of Siddis who according to the historical records have been brought as slaves by the Portuguese (Bhattacharya 1970), and the mtDNA, Y chromosome as well as autosomes coherently trace their sub-Saharan ancestry to Bantus (Shah et al. 2011; Narang et al. 2011). Makranis of Pakistan, also a part of East African slave trade show the presence of sub-Saharan mitochontrial haplogroups and 12% sharing of the Y chromosome genepools (McElreavey & Quintana-Murci 2005). Other invasions include that by Arabs followed by the Muslim legacy and ruling by Mughal emperors. However, studies on uniparental markers and autosomal STR suggest that spread of Islam was largely a cultural phenomenon with detectable traces of their lineages, primarily from Iran and Central Asia (Eaaswarkhanth et al. 2009a; Eaaswarkhanth et al. 2009b) suggesting minor gene flow.

Language shift

It is typically through the process of language assimilation, replacement or transfer that people in the language contact zone who initially speak more than one language abandon the use of their original language in favor of another with only minor or no contribution of the genes from a source population (Diamond & Bellwood 2003). The idea behind the language shift could be social and economic uplift of the minor group, or forced language adoption by ruling authorities etc. The original language however does not disappear quickly, rather requires generations to do so. One of the genetic studies on Mushars (Chaubey et al. 2008a), a population of Northern India (known to have undergone language shift from Austroasiatic to Indo-European), involving study

of maternal and paternal loci revealed haplotype sharing with both the groups albeit the sharing was higher with the AA groups. It was suggested that a change in their subsistence pattern from hunter-gatherers to agriculturists and elite dominance facilitated the language shift, perhaps because they were mostly working in the lands owned by the Indo- European speaking landlords (Chaubey et al. 2008a).

Therefore, several boundaries acting at different levels play a pivotal role in shaping the genetic diversity of South Asia. Being a home to the practice of strict endogamy thus resulting in small endogamous pockets, multi-layered social structure with unique caste system, language shifts and sex-specific admixture, all these aspects contribute majorly to the complexity of the genetic makeup of South Asians. It is interesting to note how these socio-cultural factors have also shaped the evolution of adaptive traits which have been addressed in more detail in Ref I and Ref II.

2.4.2. An insight to the genetic structuring of South Asians: from the uniparental markers to the markers of new genomic era

Social anthropologist Irawati Karve in her book (Karve 1968) described population structure of Indians as–

> "*It is like a patchwork quilt where bits of material of the same color and shape may be used in a pattern, but where each bit may be of an origin different in place and time*".

The genetic origins of South Asian populations have not been fully elucidated as there are significant gaps in our knowledge about the earliest settlers of the subcontinent. This could be partly because of the lack of archaeological data from sites that the first settlers might have occupied – many of these sites could have been in the coastal areas and by now under the water considering the rise of sea levels over the last 50,000 years. Secondly, data on genomic variation of South Asians has been underrepresented in the publicly available datasets that have been widely used to study human genetic variation on a global scale (Hinds et al. 2005; Frazer et al. 2007; Li et al. 2008), although over the recent years some attempts have been made in this direction (discussed in detail in chapter 2.4.3). Much of our current state of knowledge regarding the genetic structure of South Asians has been based on analysis of haploid genomes; maternally inherited mtDNA, paternally inherited Y chromosome (mtDNA and Y chromosome) and autosomes using high density SNP microarrays (Chaubey et al. 2007; Reich et al. 2009; Chaubey et al. 2011; Metspalu et al. 2011). In the following section, I provide an overview of the archeological evidence known till date, the genetic structure of South Asians as deduced from the uniparental and biparental markers and in the end I discuss how their inclusion in human diversity panels and in-depth study can be highly informative for understanding human genetic variation on a global scale.

Archaeological and paleoanthropological evidence till date

There have been at least two radically different models proposed about the arrival of anatomically modern humans (AMH) to South Asia. First of them suggests that the earliest settlers arrived the sub-continent dispersing from Africa *via* Arabia through the Southern coastal route along the Indian ocean rim about 40,000 to 60,000 years ago (Thangaraj et al. 2005; Macaulay et al. 2005). Therefore, it might have provided the first genetic corridor for the expansion of the humans settled in Eurasia and further expansion to different parts of Eurasia. It is also referred as a "coastal-express model" because it has been suggested that the expansion was really rapid, reaching to Australia by 50,000 years ago. This model is supported by the genetic evidence from mtDNA diversity which suggests a period between 20–40 KYA for initial settlements (Atkinson et al. 2008). A recent synthesis of genetic evidence based on mtDNA and archaeology also suggests the same (Mellars et al. 2013). The second model suggests that modern humans were present in South Asia before 75,000 years and even as early as 130,000 years based on archeological evidence (Petraglia et al. 2007) below the grey ash beds in the Jurreru valley associated with Mount Toba eruption.

There has been a glaring lacuna in the fossils or archeological evidence regarding human settlements in this part of the world which could have been important in understanding the early human settlements in the subcontinent and dispersal routes from African homeland. Nevertheless, the archeological finding till date points at the occupation of the Indian subcontinent by hominins (Paddayya 2001; Petraglia & Allchin 2007). A partial cranium (dated about 250,000–300,000 years old) found in the Narmada valley at Hathnora, Madhya Pradesh in 1982 (Sonakia & Kennedy 1985) as a part of geological and paleontological survey marks the evidence of first hominins in South Asia. The taxonomic status of the cranium still remains disputed with originally being classified as *Homo erectus* (Sonakia & Kennedy 1985), later as "archaic" *Homo sapiens* or *Homo heidelbergensis* and finally as a distinct unkown intermediate species of *Homo* named as *Homo sp indet.* (Athreya 2007; Sonakia & Kennedy 1985; Kennedy & Langstroth 2011; Sankhyan et al. 2012). The subsequent finding of two clavicles and a partial left 9^{th} rib (possibly of a female) near the same site (Sankhyan 1997; Sankhyan 1997; Sankhyan 2005) have placed Central Narmada Valley at the core of the earliest settlements during Middle Pleistocene era. A recent excavation in Netankheri, 3 km upstream of Hathnora, along the banks of the same river has brought forward yet another evidence of a femur and humerus, presently housed in the Palaeoanthropology Repository of the Anthropological Survey of India (Sankhyan 2012). It has not been yet clear if it marks the evidence of one of early modern humans as the precise dating has not been done but its stratigraphic location suggests it being below the Youngest Toba Ash i.e. older than 75 KYA. However, one of the clear cut evidence of the earliest fossils of modern humans in South Asia was found in Sri Lanka (Kennedy & Deraniyagala 1989; Deraniyagala 1984; Deraniyagala 1992), which has been dated to 28,000 to 34,000 years ago (Mellars 2006).

Additional indirect evidences are available as well; a small blade assemblage at site 55, Riwat, Northern Pakistan dated to 45 KYA (Dennell & Anwar 2004; Dennell & Anwar 2004; Dennell et al. 1988) and a microlith assemblage in the Indian subcontinent dated to 35 KYA (Clarkson et al. 2009).

mtDNA evidence

Maternally inherited mtDNA evidence suggests the existence of a number of deep autochthonous lineages (Kivisild et al. 2003; Palanichamy et al. 2004; Metspalu et al. 2004; Chaubey et al. 2007; Chaubey et al. 2008b; Thangaraj et al. 2006) in South Asia, emerging from the basal nodes of two founder lineages, macrohaplogroups M and N (including R) at 40–60 KYA (Richards et al. 2000; Sun et al. 2004; Pala et al. 2012; Palanichamy et al. 2010; Fernandes et al. 2012; Chaubey et al. 2008b). These autochthonous mtDNA haplogroups include U2a,b,c, R5-8, R30, R31, N1d ad N5, M2-6 and M30-47 (Thangaraj 2006, Chaubey 2007). Besides, the autochthonous lineages, mtDNA pools have evidence of gene flows from West and East Eurasian populations. Tibeto-Burman speakers inhabiting in North East India, demonstrate clear signals of shared ancestry with East Asian populations via mtDNA haplogroups A, B, C, D, E and G (Cordaux et al. 2003, Metspalu et al. 2004). West Eurasian mtDNAs (JT, HV, U4, U5, U7, U8, K), account for ~5% of the Indian mtDNA lineages, although the proportions vary and likely relate to the migrations during the Holocene period although their precise sources and timings remain yet to be determined. However, the pattern of mtDNA distribution does not correlate with language or caste/tribe and the deep rooted lineages are spread all over the subcontinent (Kivisild et al. 2003; Palanichamy et al. 2004; Sun et al. 2006; Thangaraj et al. 2006; Thangaraj et al. 2008; Chandrasekar et al. 2009). To summarize, the pool of mtDNA lineages found in South Asia is an amalgam of predominantly local South Asian-specific variation with limited, but nevertheless well detectable recent gene flows from West and East Eurasian sources.

Y chromosome evidence

The non-recombining haploid Y chromosome is inherited through the male line of descent. On the similar lines as that of mtDNA, South Asian paternal lineages are mainly derived from pan-Eurasian founder haplogroups C, F and K (Kivisild et al. 2003; Sengupta et al. 2006; Underhill & Kivisild 2007; Karafet et al. 2008). The most common Y chromosome haplogroups in South Asia are R1a, L, R2, H, C, J, and O2a. Of these, C5, F*, H, R2 and L1 are virtually absent elsewhere outside South Asia and are therefore likely of the indigenous origin in the subcontinent (Sengupta et al. 2006; Sahoo et al. 2006), mirroring the evidence drawn from majority of mtDNA haplogroups. On the other hand, R1a, O2a and J are haplogroups that are common in many other regions of Eurasia. R1a which is highly frequent in many populations of South Asia, Central Asia, Altay and Eastern Europe has been suggested to derive from an initial diversification somewhere in West Asia about 5,800 years ago (Underhill et al. 2014). On the other hand, phylogeography of H1a1a, prevalent among

Roma population suggesting their recent South Asian origin in Northwest India (Rai et al. 2012).

Haplogroup O2a, characterized by the marker M95, is present in high frequencies both in Indian (Munda) and SE Asian Austroasiatic (AA) speakers and provides strong evidence for the genetic link between them. On the other hand, mtDNA evidence available till date provides clear distinction between them, with Mundas sharing their basic mtDNA haplogroups with other Indian populations. An attempt to date the coalescent age of O2a-M95 using STR for Indian AA speakers, was estimated to be 8.8 ± 2.2 KYA (Zhivotovsky et al. 2004; Sengupta et al. 2006). Kayser and his colleagues (Kayser et al. 2003), although using a higher mutation rate than cited in the above study, also observed similar coalescence age for SE Asian lineages of O2a. Thus, comparing the mutation rates and ages, it was concluded that the mtDNA diversity was higher in SE Asian than among Indian AA speakers. Hence, origin of AA speakers coupled with haplogroup O2a has been enigmatic. In Ref III, we have attempted to have a closer look by using Y chromosome markers at a higher resolution and autosomal markers to see what new information they provide about the origin of these populations.

Evidence from autosomal markers

With the dawn of the genome-wide studies the picture about the genetic structure of the people of the Indian subcontinent became further refined. Two major studies (Reich et al. 2009; Metspalu et al. 2011) which involved high throughput genome-wide SNP variation analysis of Indian samples, suggested that there are two major components which make up the ancestry palette of most of the Indians living in the subcontinent (Figure 13) - one that is autochthonous to the subcontinent (referred as Ancestral South Indian (ASI) in Reich et al. 2009) and other which is closely related to West Eurasians (referred as Ancestral North Indian (ANI) in Reich et al. 2009). The degree of the ancestry proportion varies, but is evident in most of the populations studied (Reich et al. 2009), although West Eurasian component is lacking in AA and Andamanese (Reich et al. 2009; Metspalu et al. 2011). However, we do not know precisely when these two components underwent genetic admixture. An attempt to date the event (ANI-ASI admixture) by Moorjani and her colleagues (Moorjani et al. 2013) estimate it to be as recent as around 1,900 to 4,200 years ago.

To summarize, both the genetic markers, haploid as well as diploid, confirm the high genetic diversity and complexity of the South Asian genome.

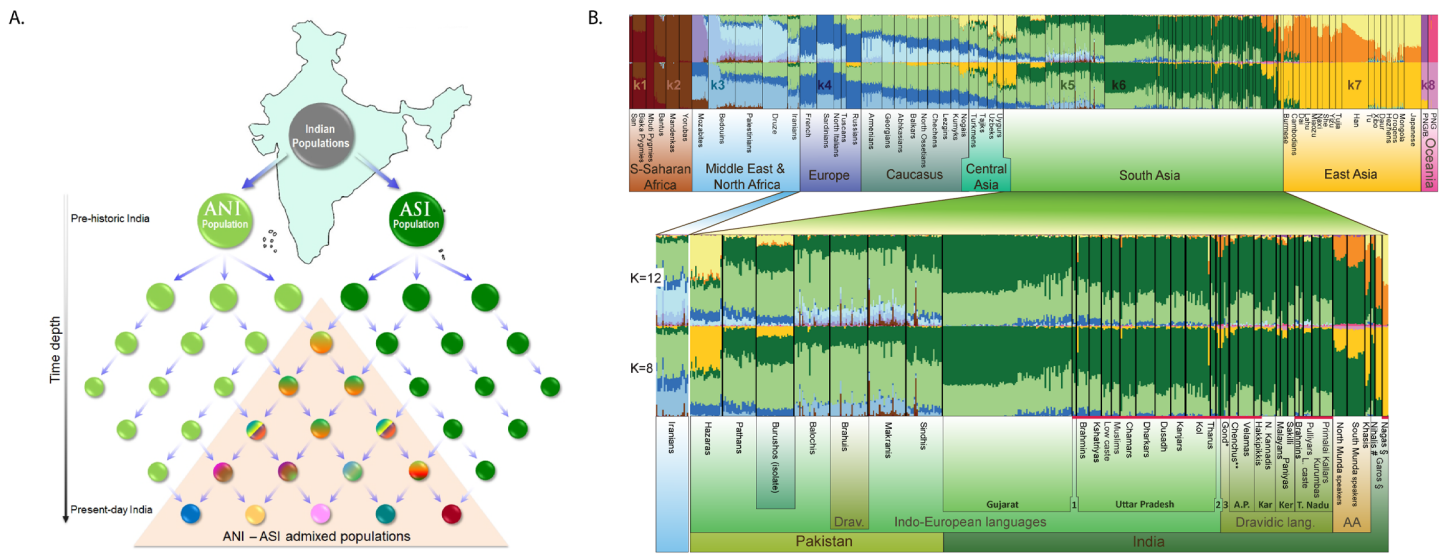

Figure 13: Two major ancestry components in South Asia. (A) Schematic diagram of the process of formation of present day Indian populations from two hypothetical ancestral populations (ANI and ASI), with Onges being an exception having only ASI ancestry component (Adapted from (Reich et al. 2009; Tamang et al. 2012). (B) ADMIXTURE plot representing genome-wide structure of Indian populations as revealed by 530,000 SNPs at K=8 and K=12 (Metspalu et al. 2011).

2.4.3. Current state-of-the-art of South Asians in existing genome diversity panels

Over the last decade, advancements in sequencing and genotyping technology have greatly revolutionized the study of human population genetics, which has expanded our knowledge about human genetic variation. Several global population sets have been genotyped using high density SNP microarrays (International HapMap Consortium 2005; Hinds et al. 2005; Li et al. 2008). This generation of population-level data has expanded from International HapMap Project focusing on 3.1 million SNPs among three major continental populations (Yoruba, Han-Chinese + Japanese and Europeans) to HGDP (Human Genome Diversity Project) dataset which included 51 worldwide populations (Li et al. 2008), further to 1000 genomes including 15 populations (1000 Genomes Project Consortium 2010). The common goal of these projects was to provide an integrated map of human genetic variation which can be further used to explore the yet unanswered questions about the human evolutionary past – in terms of population origins, migrations, ancestry proportions further extending to understanding of genetic architecture of human adaptations, patterns of selection and etiology of infectious diseases. However, South Asian populations are underrepresented in all the major datasets even including the Pan Asian Consortium representing a small fraction of South Asian diversity with 9 populations from Indian subcontinent (HUGO Pan-Asian SNP Consortium 2009). HGDP population dataset included samples from six Pakistani populations and HapMap Phase 3 included Gujarati population from Texas but it is unclear how well these groups represents the genetic diversity of the entire Indian subcontinent. The 1000 Genome Project Phase 3 aspires to be far more representative of the geographic range of major population groups of South Asia extending the data for additional six populations, including Bengalis from Bangladesh, Gujaratis (Houston, USA), Punjabi (Pakistan) and Sri Lankan Tamils, Indian Telugus (UK). However, none of these groups were sampled from India. Therefore, it is quite likely that these panels of global markers would be still missing out a substantial fraction of the indigenous genetic diversity of a large number of Indian ethnic groups that represent South Asia. Studies on expatriates have proved that they are genetically more homogeneous (Rosenberg 2006, Metspalu et al. 2011) than the ethnic populations. Moreover, the genetic architecture of a trait may be distinct for each continental region demonstrating patterns of local adaptation (Norton et al. 2007; Tishkoff et al. 2006; Jarvis et al. 2012), for example in skin pigmentation, lactase persistence or stature. Additionally differences in disease risks and protective alleles across populations (Sanghera et al. 2008; Chen et al. 2012; Corona et al. 2013) also vary, which implies an important role of population differentiation in the evolution of complex diseases. Moreover, results of selection scans performed for diseases (eg type 2 diabetes, prostrate cancer, osteoporosis) have demonstrated that there is substantial difference in their prevalence across ethnic groups (Radha & Mohan 2007; Dhandapany et al. 2009). Hence, study of

diverse populations, in particular South Asian populations, is pivotal for understanding the human genetic variation and molecular adaptation further ascribing to the fact that they are genetically the second most diverse population in the world after Africans (Xing et al. 2010; Atkinson et al. 2008).

Attempts towards Pan Indian consortium

To overcome the shortcoming and significant gaps in the global human genetic variation, one of the first attempts was a 240,000 SNP based study (Xing et al. 2009) comprising of 6 populations from 2 of the 29 states of India. In the subsequent years, understanding the need of an integrated map of South Asian genomes or in other words a Pan Indian data resource, first large scale comprehensive approach was made by Indian Genome Consortium covering critically chosen 405 genome-wide SNPs genotyped in a panel of 1871 indigenous samples from 55 populations representing all major language families (Austroasiatic, Tibeto-Burman, Indo-European, Dravidian), social levels (caste, tribe, religious group), and spanning 6 major geographical zones of Indian subcontinent (Ghosh 2005), (Brahmachari et al. 2008). Phase I of the project intended to provide a schematic extent of genetic diversity in India whereas Phase II covering 3824 SNPs from 834 candidate genes aimed at important inferences about genetic architecture of disease phenotypes and other complex traits (Narang et al. 2010). Though it did suffer from ascertainment bias and had lower coverage of SNPs, the sample coverage was wide covering the major geographical zones and the major language families. The consortium was able to come up with some novel variants for certain diseases using GWAS, but a complete genomic dissection of the Indian population has yet to be ascertained.

2.4.4. Why it is important to understand South Asian genetic variation?

Being home to about one-fifth of the world population and staggering disease burden (Ghaffar et al. 2004; Lopez et al. 2006), South Asia undoubtedly deserves the attention of genetic studies. The study of South Asia is important for our general understanding of human diversity and for answering questions related to human origins, dispersal routes from out of Africa, formation of the present-day population structure in Eurasia, effects of the spread of farming on genetic diversity and phenotypic variation, cultural and social adaptations and for the understanding the etiology of infectious diseases.

Implications in disease genetics

Individuals with South Asian ancestry have been shown to have greater susceptibility to certain diseases e.g. type 2 Diabetes (T2D), coronary artery disease (CAD), ulcerative colitis, malaria (Kumar 2012), hemoglobinopathies (Williams & Weatherall 2012). It has been observed that some of the causal variants of such diseases are population specific (Sanghera et al. 2008; Bye et

al. 2011). Furthermore, selection pressures acting on some of the genes for example *NAT2* are also population specific (Sabbagh & Darlu 2005; Sabbagh et al. 2008). The evaluation of genetic diversity where representatives of indigenous populations is crucial, GWAS of consortiums representing cases and controls of Asian Indians have been slightly lucky in identifying new variants. Genome-wide studies on South Asians living in India and Asian Indians (living in USA/UK) done till date have been fundamental in identification of novel variants related to T2D (Sanghera et al. 2008; Been et al. 2011; Kooner et al. 2011; Saxena et al. 2012), blood pressure variation, adiposity (Dwivedi et al. 2012), Ischemic stroke (Yadav et al. 2013), anaemia; CAD (Coronary Artery Disease (C4D) Genetics Consortium 2011; Mehta 2011).

The genomic era has recently undergone major developments thanks to the introduction of the next generation sequencing technologies and the drastic drop in sequencing costs. Adding a step further, with the whole genome sequences we will now be able to gain more insights about the detailed population structure and accelerate the identification of both common and rare novel variants which can be pivotal for understanding of diseases. Therefore, sampling, genotyping and sequencing of additional South Asian populations will undoubtedly continue to be important for the general understand of human diversity patterns, adaptations and diseases.

3. AIMS OF THE PRESENT STUDY

Our first general aim was to study the genetic architecture of two major human adaptive traits - skin color and lactase persistence. Our second general aim was to study from a geneticist point of view on a common and important, but hitherto not widely enough investigated chain of events in a well defined model system: migration > admixture > language change, among Austroasiatics, one of the four major language families spoken in South Asia. The specific aims as formulated in the planning of the three research papers included in my thesis are outlined below:

Ref I:

a) to quantify the skin pigmentation variation among indigenous populations of South Asia and assess the role of rs1426654 SNP;
b) to examine if the allele associated with light skin (rs1426654-A) in *SLC24A5* is identical by descent in Indian and European populations;
c) to assess the patterns of positive selection acting upon *SLC24A5* gene among the studied global populations;
d) to estimate the coalescent time of the allele (rs1426654) associated with lighter skin pigmentation.

Ref II:

a) to sequence a part of the *LCT* enhancer region, known to harbor currently known LP alleles among South Asians to study the distribution of the known and novel variants;
b) to assess how the frequency distribution of the mutations in the *LCT* enhancer region complies with the lactase persistence phenotype frequencies reported from the Indian sub-continent;
c) to study how the geographic distribution of the -13910T allele frequencies correlate with language, geography and varying subsistence patterns in India.

Ref III:

a) to use genome-wide data to test whether Indian Austroasiatic speakers have ancestry components that could relate them with Southeast Asian Austroasiatic speakers;
b) to perform a high resolution phylogeographic analysis using uniparental markers to determine sex-specific proportions of presumed Southeast Asian ancestry in Indian Austroasiatic-speaking populations;

4. MATERIAL AND METHODS

The description of the DNA samples analyzed in the present study, along with the experimental and analytical methods used, are described in detail in the respective research articles and/or their supporting materials. DNA samples were obtained from volunteers after receiving their informed consent in accordance with the guidelines of the ethical committees of the institutions involved.

Table 2: A cumulative table representing the sample sets, their location and methodology used in the three articles included in the thesis.

Study	Samples	Study /Methodology	n	Origin/Source
Ref I	Cohort A	Genotype-Phenotype association study	1228	India
	Cohort B	Genotype-Phenotype association study	446	India
	Cohort C	Genotyped for rs1426654 SNP (by Sequencing/RFLP)	1054	India/Sri Lanka
		Phylogeography of rs1426654 SNP	1446 2763	India world-wide*
	Global sample set (8 major groups)	Resequencing of *SLC24A5* (11.74kb) Tests of selection (Tests of neutrality) Coalescence age estimates	95	world-wide
	Genome-wide dataset	Tests of selection (XP-EHH, iHS)	145 890	India world-wide*
Ref II	South Asia sample set	Sequencing of enhancer region of LCT (427/706 bp)	2284	India/Nepal
	Genome-wide genotype dataset	Haplotype data analysis	199 506	India world wide*
Ref III	Sample set 1	mt DNA	578 2938	India South, East and Southeast Asia*
	Sample set 2	Y chromosome	1563 4547	India/Southeast Asia South, East and Southeast Asia*
	Sample set 3	Genotyped for 1540T/C SNP (Sequencing/RFLP)	1077	India
	Sample set 4	Autosomal data analysis (ADMIXTURE run)	41 574	India/Southeast Asia world wide*
	Sample set 5	O2a-M95 haplogroup samples (14 Y STRs)	342	South, East and Southeast Asia

* Published datasets used in the study, details can be found in the respective research articles and/or supplementary material.

5. RESULTS AND DISCUSSION

5.1. Investigating the role of *SLC24A5* in skin color variation and beyond (Ref I)

South Asia exemplifies as one of the world's major hotspots of skin color variation and therefore, serves as a perfect model to decipher the genetic underpinnings of human skin color variation and for a general understanding of it's evolutionary history. *SLC24A5,* has been important in the catalog of pigmentation genes because it explains about 25–38% skin pigmentation differences between Europeans and Africans (Lamason et al. 2005). This genetic variant has also been identified to be associated with skin color in South Asian immigrants of UK (Stokowski et al. 2007). However, it's spread pattern and significance among native populations of the Indian subcontinent has not been ascertained till date. Therefore, with this prior genetic evidence at hand, we set out to study the distribution of melanin index in India and thereafter, significance of *SLC24A5* in skin pigmentation variation of the present-day Indians. Besides this, we resequenced 11.74 kb of the *SLC24A5* gene in a global sample set, to assess the world-wide diversity and selection patterns, study the phylogenetic relationships between the populations studied and to determine the coalescent age of the light skin variant (rs1426654-A). South Asian dataset comprised of two cohorts (Cohort A and B) with skin color measurements and Cohort C with country-wide representation of 46 ethnic populations. Our global sample set used in the resequencing project comprised of 95 multiethnic individuals. This included 70 subjects from HGDP-CEPH (Human Genome Diversity Cell Line Panel) and rest from our own collections, representing 8 major geographical zones (Africa, Middle East, North/Central Asia, Europe, East Asia, South Asia, America and Melanasia).

Assessment of the skin pigmentation measures in two primary cohorts (Cohort A and B) suggested that South Asians show a wide variation in their melanin indexes (MI 28–79). We tested the association of the rs1426654 SNP with pigmentation differences between the low (MI<38) and high (MI>50) MI groups of Cohort A, a largely homogeneous sample from Andhra Pradesh of South India (Cohort A, n=1228). We found that the effect of genotype was highly significant ($p = 2.4 \times 10^{-31}$) (Figure 1 in Ref I). A replication study carried out in an ethnically and geographically heterogeneous set of 10 populations sampled across India (Cohort B, n = 446), showed similar results ($p = 3.24 \times 10^{-8}$), thereby confirming that rs1426654 SNP in the *SLC24A5* gene is significantly associated with skin pigmentation variation in South Asia. Our estimation of effect size, using multiple imputation model in Cohort A, suggested that rs1426654 SNP explains almost 27% of the variation in the studied cohort.

The detailed map of the derived variant (rs1426654-A) across 1573 individuals from 54 ethnic groups, comprising all the three cohorts of the present study (Figure 2 of Ref I), revealed that the light skin associated allele is

ubiquitous and it's frequency varies substantially across populations in South Asia (0.03 to 1). Testing the rs1426654 allele frequency patterns across various language families and geographic zones of India by Mantel test revealed that both language and geography have significant influence on its frequency ($p < 0.001$). We also observed a distinct geographic structure of the variant, with higher frequencies in the North (0.70 ± 0.18) and Northwest regions (0.87 ± 0.13) and a declining pattern as one moves further South and East.

Studies dissecting the genetic structure of South Asians have revealed the presence of two distinct ancestry components: ancestral North Indian (ANI) and ancestral South Indian (ASI) components (Reich et al. 2009), the former reflecting shared ancestry between Indian and West Eurasian populations. As the global distribution of rs1426654-A depicts higher frequencies skewed to the West Eurasian populations (Figure 2 of Ref I), reaching almost fixation in Europe, we sought to examine if there is any correlation between them. Based on synthesis of the available genome-wide SNP variation data for Indian populations from the literature (Reich et al. 2009; Chaubey et al. 2011; Yunusbayev et al. 2012; Behar et al. 2010) and the data obtained in the present study, we performed analysis by ADMIXTURE program. On comparing the proportions of the k5 light green ancestry component obtained at K=7 were with the observed allele frequencies (rs1426654-A) from the present study, we found a significant positive correlation between them ($r= 0.90$, $p<0.0001$) (Figure S2 of Ref I).

Our study of the light skin variant (rs1426654-A) carried out at the micro-geographic level, further revealed strong correlation of the allele frequencies with the demographic history of the populations. Saurashtrians, who are known to have migrated from Gujarat in the West to Madurai (Tamil Nadu) in the South, have a relatively high rs1426654-A allele frequency (0.70) compared to their local neighbors, reflecting their north-western ancestry. Similarly, populations with higher proportion of West Eurasian ancestry, like Toda from the Nilgiri Hills (Tamil Nadu, South India), demonstrate higher A allele frequency (0.86) than their geographical neighbors Kurumba (0.20). Brahmins, irrespective of their different sampling locations (North, Central or South India), possess similarly high rs1426654-A allele frequencies (0.70–0.90) across the subcontinent, which may reflect their common origin.

A phylogenetic tree reconstructed based on our resequencing data of *SLC24A5* (11.74 kb of the gene resequenced in 95 world-wide individuals) (Figure 3 of Ref I), demonstrated that rs1426654 was the only non-synonymous SNP, that defined a common and geographically wide-spread branch. Interestingly, populations showing higher frequencies of rs1426654-A allele (Middle East, Central Asia, South Asia and Europe) demonstrated low levels of intra- and inter-population diversity in *SLC24A5* (Figure 4 of Ref I) as calculated on the basis of average pairwise differences observed in the sequence data. Furthermore, our phylogenetic analyses revealed that all the carriers of the rs1426654-A allele (from Europe, sub-Saharan Africa, the Middle East, South,

North and Central Asia) share this mutation as identity-by-descent (Figure S3 of Ref I).

SLC24A5 has been well-documented in the previous studies of genome-wide scans as a candidate of positive selection among Europeans (Voight et al. 2006; Sabeti et al. 2007; Coop et al. 2009). Our selection tests were performed using the following datasets and analytical methods: a) sequence-based selection tests on resequencing data; b) haplotype-based selection tests for genome-wide genotype data (a merged dataset of Illumina Infinium 650K, 610K and 660K of 1035 individuals including 145 Indians and samples representing Africa, Middle East, Europe, Central Asia, Pakistan, East Asia and Oceania from published datasets). For our sequenced-based tests, none of the populations showed significant departures from neutrality except for Europeans (Tajima's D, p= 0.02 and Fu and Li F* (p= 0.04)). For our halpotype-based selection tests, XP-EHH (Cross-population Extended Haplotype Homozygosity) scores show evidence of positive selection among populations of Europe, Middle East, Pakistan, Central Asia and North India. Similarly, scores from our iHS analysis (Integrated Haplotype Score) had empirical p values for Central Asia and North India. It was further interesting to note that both XP-EHH and iHS scores suggested that positive selection has occurred among North Indians (within top 5% and top 1% respectively), but not among South Indian populations.

The coalescent age of rs1426654-A mutation is crucial for understanding of the evolution of light skin. Earlier attempts for age estimates of this locus, only few though, have been largely confined to on estimation of onset of selective sweep, rather than estimation of coalescence time (Voight et al. 2006; Beleza et al. 2012). Beleza (Beleza et al. 2012) and her colleagues, while focusing on the analyses based on microsatellite loci, dated selective sweep at *SLC24A5* to 11.3 KYA (95% CI 1-55.8) and 18.7 KYA (95% CI 5.8-38.3) when using additive and dominant models, respectively. Our coalescence date estimates of the rs1426654-A allele at 21 ± 10.3 KYA, using rho-based method, and, at 28 KYA (95% CI 5-58.4) using BEAST, predate the estimates of selective sweep as predicted. The difference in the age estimates (although showing substantial overlap) can be attributed to the different parameter settings and model assumptions used in the methods. The wide margins also evident in the earlier calculated sweep dates (Beleza et al. 2012), could be due to the fact that power of our analysis was limited by the need to reduce the sequence range due to high LD and also low level of sequence variation.

Main contribution to the field: Both demography and selection play an important role in skin color variation of South Asians. This aspect observed in the study, helps us to understand the different mechanisms that contribute to the global spectrum of human skin color.

5.2. Exploring the genetics of lactase persistence of Indian Dairyland (Ref II)

Ability to digest lactose in adulthood is a polymorphic phenotypic trait in humans, which allows continuous production of enzyme lactase-phlorizin hydrolase (LPH) post weaning. India is one of the largest producers of milk today (Food and Agriculture Organization of the United Nations 2009) and also is accounted for high consumption of dairy products. However, the genetic basis of lactase persistence in this region is poorly known. Previous studies have shown that this trait is useful for humans and that it has evolved multiple times, suggestive of convergent evolution (Mulcare et al. 2004; Itan et al. 2010; Gerbault et al. 2011). Therefore, it was important to study if there is a genetic evidence for independent evolution of the trait in the Indian subcontinent. To this end, we carried out a comprehensive study involving sequencing of 2284 individuals from 106 populations from South Asia, including samples from four major language families, different subsistence patterns, covering 22 of the 29 states of India, for the lactase (*LCT)* enhancer region known to harbor the lactase persistence (LP) alleles in different world-wide populations (Ingram et al. 2009; Itan et al. 2010).

Sequencing of the 700 bp of the *LCT* enhancer region among Indian samples revealed that, of the variants observed, -13910T allele (the primary determinant of LP among Europeans) has the highest frequency, with an average country-wide frequency of 10.3%. Besides the -13910T SNP, we identified additionally seven LP variants, of which 3 were novel. The second most frequent LP variant was -13779G. This mutation was earlier reported in a Somali individual (Ingram 2009). Interestingly, -13915C, reported in the present study in Toda, a pastoral population of Nilgiri Hills (South India), previously described in Muslims living in South India (Eaaswarkhanth et al. 2009b), had a polymorphism distinct from the reported Arabian LP variant (-13915 T> G) (Enattah et al. 2008). The first pilot study to report the existence of -13910T and so far the only genetic study carried out on Indians (Babu et al. 2010), in particular on two urban cohorts consisting of North (n=77) and South Indians (n=76), has also observed -13910T SNP as determinant of lactase persistence among Indians. However, unlike our study, they did not find any other polymorphism between -13830 to -14190 positions of the *LCT*.

Earlier studies based on lactase persistence phenotype data (Desai et al. 1970; Swaminathan et al. 1970; Tandon et al. 1981; Gupta et al. 2007; Reddy & Pershad 1972) point to an average frequency of 0.40 in India, whereas the predicted phenotype frequency in our samples, on the basis of the summary frequency of the allelic variants in the *LCT* enhancer region was much lower (0.197). This suggests that either we are missing a substantial portion of LP variants (they are outside of our sequence range), or previous studies have overestimated the actual phenotype frequency existing in India. Nonetheless, genotype-phenotype analyses based on interpolated phenotype (online GLAD database, http://www.ucl.ac.uk/mace-lab/resources/glad) and allele frequencies

indicate that except for few regions, the difference between genotypic and phenotypic frequency variations was not statistically significant.

As dairying culture has been strongly correlated to evolution of lactase persistence trait, we sought to compare the allele frequencies of the pastoralist and non-pastoralists populations included into our study. We found that 3 of the 5 pastoralist populations studied (Ror, Jat Muslim and Toda) account for the first, third and fourth highest frequencies of -13910T allele. Ror and Jat Muslims are pastoralist populations of North India whereas Todas reside in Nilgiri hills (South India). Interestingly, Toda, also demonstrated the highest allelic diversity, with the presence of 3 of the 7 different variants detected in the *LCT* enhancer region. This further accentuates the fact that dairying has been one of the major selective forces in the distribution of lactase persistence in the subcontinent consistent with previous studies (Enattah et al. 2008; Ingram et al. 2009). The phylogeography of the SNP (Figure 1 of Ref II) mapped on the basis of allele frequencies showed a general North-South and West-East cline. Comparing the allele frequencies across major language families and geographic zones, showed a significant correlation with language ($p= 1.74 \times 10^{-6}$) and geography ($p= 1.853 \times 10^{-6}$). Therefore, the geospatial patterns of -13910T variant mapped across the Indian subcontinent, reflects a striking example of interaction between subsistence strategy, language and geography (Figure 1 of Ref II).

Given the fact that we observed -13910T as the predominant segregating polymorphism of lactase persistence among the individuals studied, one of the key questions was if the Indian -13910T is identical by descent to already well known -13910T European variant. To answer this, we used the genome-wide genotype data of 705 individuals, including 506 individuals from CEPH-HGDP and 199 Indians and retrieved SNP data for the 5 Mb region surrounding -13910T locus. Analysis of a 60 kb region containing 15 SNPs, allowed us to identify 21 distinct haplotypes. We found that all the -13910T alleles in our South Asian dataset are found on the previously defined European haplotype A (Hollox et al. 2001; Poulter et al. 2003), thereby suggesting that both Europeans and Indians share the mutation as identity-by-descent. Our neighbor-joining network (Figure 2 of Ref II) reconstructed using the above genotype data clearly recapitulated similar high frequency haplotypes (A*T, A, B, C) as seen in the previous studies (Poulter et al. 2003; Bersaglieri et al. 2004).

LCT has been marked as one of the hotspots for recent positive selection among Europeans (Bersaglieri et al. 2004) which has been also recapitulated by several genome-wide scans (Sabeti et al. 2007; Voight et al. 2006; Akey 2009). Our EHH (Extended Haplotype Homozygosity) scores, calculated for long core haplotypes among four continental groups (European, Near Eastern, Pakistani and Indian populations) using program Sweep, showed that all these populations were associated with the same long range haplotype, extending up to 1Mb. However, the decay patterns were not identical across all populations. All the samples from Europe, Middle East and Pakistan were characterized by a

single steep drop of 800kb from the core region, whereas the Indian samples exhibited a marked ladder-like pattern of decay (Figure 3 of Ref II).

Main contribution to the field: Provides the first comprehensive genomic landscape of the enhancer region of *LCT* among South Asians.

5.3. Austroasiatics: Who are they? Where did they come from? (Ref III)

Austroasiatic (AA) speakers constitute the eighth largest language family in terms of native speakers in the world (Lewis 2009). Nevertheless, their origin and path of dispersal to their present dwelling in South and Southeast Asia remains disputed. Inferences from the mtDNA and Y chromosome studies till date (Basu et al. 2003, Metspalu et al. 2004; Cordaux & Stoneking 2003; Kumar V. et al. 2007; Sengupta et al. 2006; Kayser et al. 2006), point towards two contrasting models. The first of these places the origin of AA speakers in Southeast Asia with a later dispersal to South Asia during the Neolithic, whereas the second hypothesis advocates pre-Neolithic origins and dispersal of this language family to Southeast Asia from South Asia. Here, we used in parallel, mtDNA, Y chromosome and most importantly, autosomal markers to gain a fine-grained understanding of the genetic structure of the AA speakers living in South Asia. In total, 45 samples from India, including (22 AA speakers (both from Munda and Khasi-Aslian speakers), 4 Tibeto-Burman and 19 Dravidian (Behar et al. 2010)) were genotyped for 650,000 SNPs using Ilumina 610K. Subsequently, these results were combined with global dataset (Li et al. 2008) which included 10 samples from Cambodia, as a representation of Southeast Asian AA speakers.

Comparing the ancestry proportions of the Indian AA speakers with the other non-Austroasiatic populations of India (Figure 3B of Ref III), as revealed through the fine scale population structuring from the ADMIXTURE analysis (K=7), demonstrated the presence of a substantial proportion of a distinct East/Southeast Asian component, which was otherwise absent among non-Austroasiatic populations of India. Furthermore, we observed differences in the genetic structuring of two Indian AA speaking tribes, Munda and Khasi, complementing their linguistic and geographic distinction. Munda individuals, included in our dataset, are majorly spread over Central and East India, and speak Mundari language, whereas Khasi, who speak mostly Khasi-Aslian language, are mainly restricted to Meghalaya state of North-East India.

Earlier genetic evidence based on mtDNA, suggested absence of any South-East or East Asian ancestry signal among Indian Munda speakers (Basu et al. 2003; Metspalu et al. 2004). Conversely, we observed a notable presence (23%, SD 5%) of Southeast Asian signal among Indian Munda speakers in the ADMIXTURE plot generated from the autosomal loci (Figure 3B of Ref III). On the other hand, we also detected traces of South Asian genetic admixture among the Cambodians (16%, SD 5%). Thereby, the presence of genetic

admixture among Southeast Asian and Indian AA speakers points to bidirectional geneflow(s) alongside of the Bay-of-Bengal among the AA speakers.

Interestingly, our analyses also showed that the genetic component autochthonous to India (also referred as Ancestral South Indian in (Reich et al. 2009) was higher (almost 75%) in Munda speakers compared to Khasi (Figure 3B of Ref III). Our principal component analysis (PCA) of Indian populations in the context of other Eurasian populations carried out using a matrix of 615 samples with 189,533 SNPs, recapitulated similar pattern. Namely, our PC plot (Figure 3A of Ref III), demonstrated that Munda speakers are more closely related to Indian Dravidian speakers, whereas Khasi-Aslian and Tibeto-Burman groups from India and Southeast Asia are more similar to each other, although the Indian Khasi-Aslian also showed high affinity to Munda speakers.

Assessment of the EDAR polymorphism rs3827760 (1540C), a locus of positive selection among East Asians (Kamberov et al. 2013) allowed us to determine the relative proportions of East Asian-specific ancestry among the Indian AA populations studied. Country-wide screening of the variant among 1077 individuals including 49 Indian populations of the sub-continent showed that this variant at the highest frequencies (61%) in Tibeto-Burman speakers, consistent with their predominantly East Asian ancestry inferred from autosomal and uniparental loci. Meanwhile, the AA populations, Munda (0.05) and Khasi-Aslian (0.40) showed contrasting allele frequency patterns. Notably, this variant was very low or virtually absent among the Indo-European and Dravidian speakers (Figure 4 of Ref III).

Our analyses of Y chromosome involving genotyping of 12 SNPs in 553 individuals representing 13 diverse AA populations of India revealed the presence of eight distinct haplogroups. The most frequent haplogroup was O2a, which is consistent with earlier studies (Basu et al. 2003; Metspalu et al. 2004; Sengupta 2009; Sahoo et al. 2006). The microsatellite variance associated with O2a lineages was found to be higher in SE Asians than in Indians, showing overall a sharp clinal pattern from East to West (Figure 5 of Ref III) and suggesting SE Asian origin of haplogroup O2a.

Our next attempt was to date the coalescence time of M95, the characteristic marker of O2a, which provides the signal for shared Southeast Asian ancestry. Previous Y chromosome-based evidence had provided controversial dates for the shared O2a lineage either because of different sampling or genotyping approaches (Sengupta et al. 2006; Kumar V. et al. 2007). In the present study, we used 14 STR loci to provide an age estimate using individuals both from India as well as Southeast Asia to avoid any sample bias. We found that the overall age of all Y chromosomes from India and Southeast Asia carrying the M95 mutation was ~ 20 (±2.7) KYA and 15.9 (± 1.6) KYA for M95 diversity in India. Nonetheless, our estimate for the Indian dataset should not be taken as a genetic estimate of dispersal time of AA to India, but rather can be accounted for the upper boundary for any dispersal event in conjunction with the O2a lineage.

To sum up, synthesis of mtDNA, Y chromosome and autosomal analyses suggests that the present-day AA speakers, at least in case of Munda speakers, have been derived from recent dispersal(s) from Southeast Asia followed by extensive admixture with local Indian populations. The strongest signal of Southeast Asian genetic ancestry among Indian Austroasiatic speakers is maintained in their Y chromosomes, with approximately two-thirds falling into haplogroup O2a. Furthermore, the fact that Southeast Asian ancestry signal observed in Munda speakers evident in their autosomes and Y chromosome and not in mtDNA suggests two possible scenarios; either the gene flow from Southeast Asia to South Asia was largely asymmetric and gender-specific, being restricted to males, or the Southeast Asian ancestry signal in the mtDNA lineages has been lost due to extensive admixture with local populations.

Main contribution to the field: This study helps us to understand the role of gender-specific demographic processes in the peopling of the Indian sub-continent.

6. CONCLUSIONS

- Quantitative assessment of skin color among ethnic South Asians shows a wide variation in their melanin indexes, reflecting their pigmentation diversity. Significant association of rs1426654 SNP with skin pigmentation, accounting for 27% variation in the studied cohort, suggests that this SNP has been pivotal in shaping the pigmentation variation among South Asians.
- Phylogenetic analyses based on *SLC24A5* resequencing data, reveal that the light skin associated allele of *SLC24A5* (rs1426654-A) in South Asians and Europeans are identical by descent.
- Study of selection patterns across *SLC24A5*, based on our resequencing and genome-wide data, both confirm an earlier evidence for positive selection among Europeans. In addition, haplotype-based selection tests based on genome-wide genotype data, show signatures of positive selection among populations of Middle East, Central Asia, Pakistan and North India, but not in South Indians.
- Sequencing of the enhancer region of lactase gene (*LCT*) among Indian populations demonstrates that out of the 8 variants identified, -13910T has the highest frequency and widest distribution. Haplotype analyses further reveal shared origin of -13910T SNP among Europeans and Indians, associated with the same >1Mb extended haplotype.
- Country-wide frequency patterns of variants, associated with light skin (rs1426654-A) and lactase persistence (-13910T), studied among Indian populations, demonstrate that both these adaptive traits have been shaped by a combination of processes involving selection and demographic history of the populations.
- Presence of detectable Southeast Asian and South Asian ancestry components among Indian and Southeast Asian Austroasiatic-speaking populations, as revealed by their autosomal loci analyses, provides evidence of bidirectional gene flows alongside the Bay-of-Bengal.
- Genetic evidence obtained from the conjunction of mtDNA, Y chromosome and autosomal analyses of Indian Austroasiatic speakers, is consistent with the model, implying their recent dispersal(s) from Southeast Asia, followed by admixture with local Indian populations. A notable presence of Southeast Asian genetic component in the autosomes and Y chromosome of Indian Munda speakers, but not in their mtDNA, further suggests gender asymmetry in such gene flows and sex-specific admixture in the peopling of the subcontinent.

7. SUMMARY IN ESTONIAN

Adaptatiivsete tunnuste geneetika ja soo-spetsiifilised demograafilised protsessid Lõuna-Aasia populatsioonides

Inimkonna varieeruvuse oluliseks aspektiks on meie kui liigi fenotüübiline mitmekesisus: millised me välja näeme, mida me sööme, milliseid kliimatingimusi me talume ja millised on meie haigusriskid. Need fenotüübilised omadused peegeldavad inimese kohastumisi erinevate loodusliku valiku survetega minevikus, nagu muutused keskkonnas, eluviisis ja toitumisharjumustes. Nende kohastumuste mehhanismide mõistmine on teadlastele tõsiseks väljakutseks, kuna probleemistik haarab ja põimib kokku nii keskkonnast tulenevaid kui ka pärilikke tegureid ning eelajaloolisi demograafilisi protsesse. Uurimismeetodite areng on täna teinud võimalikuks loodusliku valiku jälgede otsimise ja tuvastamise genoomides. See avab tee pilguheiduks inimese evolutsioonilisse minevikku: kuidas ja millal me omandasime unikaalsed kohastumused, kuidas need mõjutasid meie bioloogilist kohasust ning millised loodusliku valiku mõjud on olnud olulised kujundamaks meid sellisteks, nagu me tänapäeval oleme. Teiselt poolt aitab erinevate populatsioonide geneetilise struktureerituse uurimine mõista demograafiliste protsesside rolli inimkonna geneetilise varieeruvuse tekkes, samuti avastada võimalikke keelevahetusi uuritavais populatsioonides. Tuleb lisada, et sellelaadsed uuringud on maailmas alles algamas.

Käesolev väitekiri keskendub kahele kohastumuse näitele, milleks on nahavärv ja laktoosi taluvus. Paljud eelnevad uuringud viitavad sellele, et nahatoonide ülemaailmne mitmekesisus on õrna tasakaalu tulemuseks selle vahel, et nahk oleks piisavalt hele, võimaldamaks vajalikul hulgal D-vitamiini tootmist madala UV-kiirgusega aladel, ja samas piisavalt tume, et pakkuda kõrge UV-kiirgusega aladel kaitset päikesepõletuse ja folaadi lagunemise eest. Senised geneetilised andmed on näidanud, et nahavärvi varieeruvusega assotsieeruvad muudatused 170 geenis. Teiseks uurisime fenotüüpi, mis kujunes välja seoses inimkonna olulise osa üleminekuga küttimiselt-koriluselt põllupidamisele-loomakasvatusele. Osana inimese kohastumisest seoses karjakasvatusega arenes evolutsioonis laktoosi taluvus – geneetiline muutus, mis võimaldab täiskasvanueas ilma terviseprobleemideta piima seedida. On tõenäone, et mõlemad kohastumused – hele nahk ja laktoosi taluvus – on inimpopulatsioonides tekkinud rohkem kui üks kord, mis näitab nende tähtsust inimese evolutsioonis. Me kasutasime mõlema tunnuse uurimiseks kandidaatgeenide meetodit. Keskendusime eelkõige Lõuna-Aasia populatsioonidele, kelle seas nahatooni varieeruvus on suur ning kus elab populatsioone erinevate piimajoomise kommete ja karjakasvatuse ajalooga. Meie integreeritud populatsioonigeneetiline lähenemine koos võrdlusandmete kaasamisega maailma populatsioonidest võimaldas leida vastuseid küsimustele kandidaat-SNP-de fülogeograafilise leviku, tekkeaja hinnangute, fülogeneetiliste analüüside ja valikumustrite kohta ning pakkuda uut vaadet nende evolutsioonilisele ajaloole Euraasias.

Väitekirja teine osa on pühendatud ühele Lõuna-Aasia asustamise kesksele küsimusele, käsitledes austroaasia keeli rääkivaid populatsioone, kes elavad nii Lõuna- kui Kagu-Aasias. Me kasutasime ülegenoomset analüüsi koos kõrge lahutusastmega genotüpiseeritud klassikaliste ema- ja isaliini markeritega selleks, et selgitada kuidas demograafilised protsessid on kujundanud austroaasia keeli rääkivate populatsioonide geneetilist struktuuri, sealjuures sõltuvana geenivoogude soolisest tasakaalust.

Käesoleva uurimuse peamised järeldused on järgmised:

- Selles uuringus mõõdeti esmakordselt kvantitatiivselt naha pigmentatsiooni varieeruvust India põlispopulatsioonide hulgas. Uuringu tulemusena selgus nende populatsioonide melaniini-indeksite suur varieeruvus (MI 28–78), mis peegeldab pigmentatsiooni fenotüübilist mitmekesisust Lõuna-Aasias. Statistiliselt oluline seos melaniini-indeksi varieeruvuse ja rs1426654-A alleeli vahel kinnitab, et *SLC24A5* geen on üks põhilisi pigmentatsiooni mitmekesisuse määrajaid Lõuna-Aasias;
- Fülogeneetilised analüüsid ja *SLC24A5* ülemaailmse varieeruvusmustri analüüs meie poolt resekveneeritud populatsioonides osundavad, et kõik populatsioonid (Euroopast, Lähis-Idast, Kesk- ja Lõuna-Aasiast), kellel on heleda nahavärvusega kaasnev mutatsioon rs1426654-A, on selle pärinud 21–28 tuhat aastat tagasi elanud ühiselt eellaselt;
- Pigmentatsiooni taseme muutuse haplotüübipõhise valiku testid maailma populatsioonide seas on kinnitanud fenotüübi positiivset valikut eurooplaste hulgas, kuid uudse tulemusena näitavad positiivset valikut ka Lähis-Ida, Kesk-Aasia, Pakistani ja Põhja-India, aga mitte Lõuna-India populatsioonide seas;
- *LCT* võimendajaregiooni sekveneerimine Lõuna-Aasia piimatarbimise diasporaa geneetika selgitamiseks näitab, et kaheksast leitud mutatsioonist (millest 3 on kirjeldatud esmakordselt) on -13910T (varasemast teadaolevalt seotud laktaasi persistentsusega eurooplaste hulgas) põhiline, selle keskmine sagedus regioonis on 10,3%. Veelgi enam, meie poolt läbiviidud haplotüübi rekonstruktsiooni analüüs osutas, et kõnealune mutatsioon on eurooplastel ja indialastel samatekkelise päritoluga;
- Heleda naha (rs1426654-A) ja laktoosi taluvuse (-13910T) mutatsioonide alleelisageduste kaardistamine India populatsioonides võimaldas selgitada varieerumise mustreid, mis peegeldavad loodusliku valiku (seoses UV-kiirguse ja karjakasvatusega) ja populatsioonide demograafilise ajaloo vastastikust mõju. Ülaltoodud aspekt, mis ilmnes mõlemas uuringus, aitab meil mõista erinevaid mehhanisme, mis kujundavad nende tunnuste ülemaailmse varieeruvuse mustreid;
- Nii keel kui ka geograafia selgitavad olulist osa laktoosi taluvusega seotud alleeli varieeruvusest. Kõrge alleelisagedus esineb karjakasvatajate populatsioonides, peegeldades keele, geenide ja kultuuri ühist levimist;

- Autosoomsete lookuste kasutamine koos mtDNA ja Y-kromosoomiga täpsustab meie varasemat arusaamist austroaasia keeli kõnelevate populatsioonide geneetilisest struktureeritusest. Neis lookustes leitud Kagu-Aasia ja Lõuna-Aasia päritolu komponentide osakaalud annavad tunnistust kahesuunalistest geenivoogudest Bengali lahe lääne- ja idapoolse osa vahel. See järeldus vastandub varasemale arusaamisele Lõuna-Aasia asustamisest, mis on peamiselt piirdunud ühesuunalise geenivoo eeldamisel;
- mtDNA ja Y-kromosoomi vastanduvad mustrid, kus mtDNA liinid on geneetiliselt lähedased teiste India populatsioonidega ja Y-kromosoomi liinid näitavad India austroaasia keeli rääkivate mundade ühist põlvnemist kagu-aasialastega, peegeldavad Kagu-Aasiast lähtunud geenivoogude soolist asümmeetriat – neid vahendasid peamiselt mehed ja neile järgnes soospetsiifiline segunemine kohalike indialastega. Me tuvastasime ka erineva geneetilise struktuuri India austroaasia keelte rääkijate hulgas, eriti mundade ja khaside vahel, mis on kooskõlas nende geograafilise ja lingvistilise eraldatusega.

Kokkuvõttes näitab käesolev väitekiri fenotüübiliselt erinevate populatsioonide paralleelse uurimise vajadust inimese geneetilise mitmekesisuse ja kohastumuste mõistmiseks. Saadud tulemused on aidanud tuua täiendavat selgust inimese nahavärvi ja laktoosi taluvuse tekkes ja levikus Lõuna-Aasias ja Euraasias tervikuna ning soo-spetsiifilisete demograafiliste protsesside kohta, mis on kujundanud tänapäeva lõuna-aasialaste geneetilist struktuuri.

REFERENCES

1000 Genomes Project Consortium, 2010. A map of human genome variation from population-scale sequencing. *Nature*, 467(7319), pp. 1061–1073.

Akey, J.M., 2009. Constructing genomic maps of positive selection in humans: Where do we go from here? *Genome research*, 19(5), pp. 711–722.

Akey, J.M. et al., 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome research*, 12(12), pp. 1805–1814.

Akey, J.M. et al., 2004. Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS biology*, 2(10), p. e286.

Allchin, F. & Allchin, B., 1974. Some new thoughts on Indian cattle. *South Asian archaeology*.

Allen, H.L. et al., 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 467(7317), pp. 832–838.

Allen Orr, H., 1999. The evolutionary genetics of adaptation: a simulation study. *Genetical research*, 74(03), pp. 207–214.

Alm, L., 1982. Effect of fermentation on lactose, glucose, and galactose content in milk and suitability of fermented milk products for lactose intolerant individuals. *Journal of Dairy Science*, 65(3), pp. 346–352.

Altshuler, D. et al., 2005. A haplotype map of the human genome.

Andrés, A.M. et al., 2009. Targets of balancing selection in the human genome. *Molecular biology and evolution*, 26(12), pp. 2755–2764.

Aoki, K., 2002. Sexual selection as a cause of human skin colour variation: Darwin's hypothesis revisited. *Annals of human biology*, 29(6), pp. 589–608.

Aoki, K., 2001. Theoretical and empirical aspects of gene–culture coevolution. *Theoretical population biology*, 59(4), pp. 253–261.

Aoki, K., Feldman, M.W. & Kerr, B., 2001. Models of sexual selection on a quantitative genetic trait when preference is acquired by sexual imprinting. *Evolution*, 55(1), pp. 25–32.

Asthana, S. et al., 2007. Widely distributed noncoding purifying selection in the human genome. *Proceedings of the National Academy of Sciences*, 104(30), pp. 12410–12415.

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.

Atkinson, Q.D., Gray, R.D. & Drummond, A.J., 2008. mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. *Molecular biology and evolution*, 25(2), pp. 468–474.

Babu, J. et al., 2010. Frequency of lactose malabsorption among healthy southern and northern Indian populations by genetic analysis and lactose hydrogen breath and tolerance tests. *The American journal of clinical nutrition*, 91(1), pp. 140–146.

Balaresque, P.L., Ballereau, S.J. & Jobling, M.A., 2007. Challenges in human genetic diversity: demographic history and adaptation. *Human molecular genetics*, 16(R2), pp. R134–R139.

Bamshad, M. et al., 2001. Genetic evidence on the origins of Indian caste populations. *Genome research*, 11(6), pp. 994–1004.

Bamshad, M.J. et al., 2003. Human population genetic structure and inference of group membership. *The American Journal of Human Genetics*, 72(3), pp. 578–589.

Barreiro, A. et al., 2008. Subnanometer motion of cargoes driven by thermal gradients along carbon nanotubes. *Science*, 320(5877), pp. 775–778.
Barreiro, L.B. & Quintana-Murci, L., 2009. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nature Reviews Genetics*, 11(1), pp. 17–30.
Barrett, R.D. & Schluter, D., 2008. Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1), pp. 38–44.
Basu, A. et al., 2003. Ethnic India: a genomic view, with special reference to peopling and structure. *Genome research*, 13(10), pp. 2277–2290.
Bayless, T., 1971. Lactase defeciency and intolerance to milk. *Viewpoints on Digestive Diseases*, 3(2), pp. 1–4.
Been, L.F. et al., 2011. Variants in KCNQ1 increase type II diabetes susceptibility in South Asians: a study of 3,310 subjects from India and the US. *BMC medical genetics*, 12(1), p.18.
Behar, D.M. et al., 2010. The genome-wide structure of the Jewish people. *Nature*, 466(7303), pp. 238–242.
Beja-Pereira, A. et al., 2003. Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nature genetics*, 35(4), pp. 311–313.
Beleza, S. et al., 2012. The timing of pigmentation lightening in Europeans. *Molecular Biology and Evolution*.
Beleza, S. et al., 2013. Genetic Architecture of Skin and Eye Color in an African-European Admixed Population. *PLoS genetics*, 9(3), p.e1003372.
Van den Berghe, P.L. & Frost, P., 1986. Skin color preference, sexual dimorphism and sexual selection: A case of gene culture co-evolution?*. *Ethnic and Racial Studies*, 9(1), pp. 87–113.
Bersaglieri, T. et al., 2004. Genetic signatures of strong recent positive selection at the lactase gene. *The American Journal of Human Genetics*, 74(6), pp. 1111–1120.
Bhasin, V., 2011. Pastoralists of Himalayas. *Journal of Human Ecology*, 33(3), pp. 147–177.
Bhatia, G. et al., 2013. Estimating and interpreting FST: The impact of rare variants. *Genome research*, 23(9), pp. 1514–1521.
Bhattacharya, D., 1970. Indians of African origin. *Cahiers d'études africaines*, pp. 579–582.
Biasutti, R., 1967. Razze e popoli della terra, vol. IV. *Torino: UTET*.
Bigham, A. et al., 2010. Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS genetics*, 6(9), p.e1001116.
Bigham, A.W. et al., 2013. Andean and Tibetan patterns of adaptation to high altitude. *American Journal of Human Biology*, 25(2), pp. 190–197.
Bigham, A.W. et al., 2009. Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. *Human genomics*, 4(2), p.79.
Biller, H. & Grand, R.J., 1990. Lactose intolerance. *Annual review of Medicine*, 41(1), pp. 141–148.
Bittles, A.H., Coble, J.M. & Rao, N.A., 1993. Trends in consanguineous marriage in Karnataka, South India, 1980–89. *Journal of biosocial science*, 25(01), pp. 111–116.
Black, M. et al., 2006. Combining genetics and population history in the study of ethnic diversity in the People's Republic of China. *Human biology*, 78(3), pp. 277–293.
Bocquet-Appel, J.-P. & Bar-Yosef, O., 2008. *The Neolithic demographic transition and its consequences*, Springer.

Bollongino, R. et al., 2006. Early history of European domestic cattle as revealed by ancient DNA. *Biology letters*, 2(1), pp. 155–159.

Boll, W., Wagner, P. & Mantei, N., 1991. Structure of the chromosomal gene and cDNAs coding for lactase-phlorizin hydrolase in humans with adult-type hypolactasia or persistence of lactase. *American journal of human genetics*, 48(5), p.889.

Bouakaze, C. et al., 2009. Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. *International journal of legal medicine*, 123(4), pp. 315–325.

Bower, C. & Stanley, F.J., 1989. Dietary folate as a risk factor for neural-tube defects: evidence from a case-control study in Western Australia. *The Medical Journal of Australia*, 150(11), pp. 613–619.

Bradley, D.G. & Magee, D.A., 2006. Genetics and the origins of domestic cattle. *Documenting domestication: new genetic and archaeological paradigms*, pp. 317–328.

Brahmachari, S.K. et al., 2008. Genetic landscape of the people of India: a canvas for disease gene exploration. *Journal of genetics*, 87(1), pp. 3–20.

Branda, R.F. & Eaton, J.W., 1978. Skin color and nutrient photolysis: an evolutionary hypothesis. *Science*, 201(4356), pp. 625–626.

Breton, G. et al., 2014. Lactase Persistence Alleles Reveal Partial East African Ancestry of Southern African Khoe Pastoralists. *Current Biology*, 24(8), pp. 852–858.

Bulhões, A. et al., 2007. Correlation between lactose absorption and the C/T-13910 and G/A-22018 mutations of the lactase-phlorizin hydrolase (LCT) gene in adult-type hypolactasia. *Brazilian Journal of Medical and Biological Research*, 40(11), pp. 1441–1446.

Burger, J. et al., 2007. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proceedings of the National Academy of Sciences*, 104(10), pp. 3736–3741.

Butler, J., 2014. Genomics and Forensic DNA Analysis. In Plant and Animal Genome XXII Conference. Plant and Animal Genome.

Byard, P.J., 1981. Quantitative genetics of human skin color. *American Journal of Physical Anthropology*, 24(S2), pp. 123–137.

Bye, H. et al., 2011. Population-specific genetic associations with oesophageal squamous cell carcinoma in South Africa. *Carcinogenesis*, 32(12), pp. 1855–1861.

Cabanac, M. & Caputa, M., 1979. Natural selective cooling of the human brain: evidence of its occurrence and magnitude. *The Journal of physiology*, 286(1), pp. 255–264.

Candille, S.I. et al., 2012. Genome-Wide Association Studies of Quantitatively Measured Skin, Hair, and Eye Pigmentation in Four European Populations. *PloS one*, 7(10), p.e48294.

Canfield, V.A. et al., 2013. Molecular phylogeography of a human autosomal skin color locus under natural selection. *G3: Genes| Genomes| Genetics*, 3(11), pp. 2059–2067.

Carto, S.L. et al., 2009. Out of Africa and into an ice age: on the role of global climate change in the late Pleistocene migration of early modern humans out of Africa. *Journal of Human Evolution*, 56(2), pp. 139–151.

Chandrasekar, A. et al., 2009. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PloS one*, 4(10), p.e7447.

Chaplin, G., 2004. Geographic distribution of environmental factors influencing human skin coloration. *American journal of physical anthropology*, 125(3), pp. 292–302.

Chaubey, G., Metspalu, M., et al., 2008a. Language shift by indigenous population: a model genetic study in South Asia. *International Journal of Human Genetics*, 8(1/2), p.41.
Chaubey, G. et al., 2007. Peopling of South Asia: investigating the caste–tribe continuum in India. *Bioessays*, 29(1), pp. 91–100.
Chaubey, G. et al., 2007. Peopling of South Asia: investigating the caste-tribe continuum in India. *Bioessays*, 29(1), pp. 91–100.
Chaubey, G., Karmin, M., et al., 2008b. Phylogeography of mtDNA haplogroup R7 in the Indian peninsula. *BMC evolutionary biology*, 8(1), p.227.
Chaubey, G. et al., 2011. Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Molecular biology and evolution*, 28(2), pp. 1013–1024.
Chen, H. et al., 2012. TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. *Nature immunology*, 13(7), pp. 691–700.
Chen, S. et al., 2010. Zebu cattle are an exclusive legacy of the South Asia Neolithic. *Molecular biology and evolution*, 27(1), pp. 1–6.
Claes, P. et al., 2014. Modeling 3D Facial Shape from DNA. *PLoS genetics*, 10(3), p.e1004224.
Clarkson, C. et al., 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(320), pp. 326–348.
Clarys, P. et al., 2000. Skin color measurements: comparison between three instruments: the Chromameter®, the DermaSpectrometer® and the Mexameter®. *Skin Research and Technology*, 6(4), pp. 230–238.
Coelho, M. et al., 2005. Microsatellite variation and evolution of human lactase persistence. *Human genetics*, 117(4), pp. 329–339.
Coelho, M. et al., 2009. On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC evolutionary biology*, 9(1), p.80.
Comas, I. et al., 2013. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nature genetics*, 45(10), pp. 1176–1182.
Cook, G., 1978. Did persistence of intestinal lactase into adult life originate on the Arabian peninsula? *Man*, pp. 418–427.
Cook, G. & Al-Torki, M., 1975. High intestinal lactase concentrations in adult Arbs in Saudi Arabia. *British medical journal*, 3(5976), p.135.
Coop, G. et al., 2009. The role of geography in human adaptation. *PLoS genetics*, 5(6), p.e1000500.
Cordaux, R. et al., 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *European Journal of Human Genetics*, 11(3), pp. 253–264.
Cordaux, R. & Stoneking, M., 2003. South Asia, the Andamanese, and the genetic evidence for an “early” human dispersal out of Africa. *American journal of human genetics*, 72(6), p.1586.
Corona, E. et al., 2013. Analysis of the genetic basis of disease in the context of worldwide human relationships and migration. *PLoS genetics*, 9(5), p.e1003447.
Coronary Artery Disease (C4D) Genetics Consortium, 2011. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nature genetics*, 43(4), pp. 339–344.

Cosentino, M.J., Pakyz, R.E. & Fried, J., 1990. Pyrimethamine: an approach to the development of a male contraceptive. *Proceedings of the National Academy of Sciences*, 87(4), pp. 1431–1435.
Darwin, C., 1871. Sexual selection and the descent of man. *Murray, London*.
Darwin, C., 2004. *The descent of man*, Digireads. com Publishing.
Dembinski, G.M. & Picard, C.J., 2014. Evaluation of the IrisPlex DNA-based eye color prediction assay in a United States population. *Forensic Science International: Genetics*, 9, pp. 111–117.
Dennell, R. & Anwar, M., 2004. *Early hominin landscapes in northern Pakistan: investigations in the Pabbi Hills*, British Archaeological Reports Ltd.
Dennell, R., Rendell, H. & Hailwood, E., 1988. Late Pliocene artefacts from northern Pakistan. *Current Anthropology*, pp. 495–498.
Deraniyagala, S.U., 1984. Mesolithic stone tool technology at 28,000 BP in Sri Lanka. *Ancient Ceylon*, 5, pp. 105–108.
Deraniyagala, S.U., 1992. *The prehistory of Sri Lanka: an ecological perspective*, Department of archaeological survey, government of Sri Lanka.
Desai, H. et al., 1970. Incidence of lactase deficiency in control subjects from India. Role of hereditary factors. *Indian journal of medical sciences*, 24, pp. 729–736.
Dhandapany, P.S. et al., 2009. A common MYBPC3 (cardiac myosin binding protein C) variant associated with cardiomyopathies in South Asia. *Nature genetics*, 41(2), pp. 187–191.
Diamond, J. & Bellwood, P., 2003. Farmers and their languages: the first expansions. *Science*, 300(5619), pp. 597–603.
Diffey, B., Oliver, R. & Farr, P., 1984. A portable instrument for quantifying erythema induced by ultraviolet radiation. *British Journal of Dermatology*, 111(6), pp. 663–672.
Dubouloz, J., 2008. Impacts of the Neolithic demographic transition on Linear Pottery Culture settlement. In *The Neolithic demographic transition and its consequences*. Springer, pp. 207–235.
Durso, D.F. et al., 2014. Association of genetic variants with self-assessed color categories in brazilians. *PloS one*, 9(1), p.e83926.
Dwivedi, O.P. et al., 2012. Strong influence of variants near MC4R on adiposity in children and adults: a cross-sectional study in Indian population. *Journal of human genetics*, 58(1), pp. 27–32.
Eaaswarkhanth, M., Dubey, B., et al., 2009. Diverse genetic origin of Indian Muslims: evidence from autosomal STR loci. *Journal of human genetics*, 54(6), pp. 340–348.
Eaaswarkhanth, M., Haque, I., et al., 2009. Traces of sub-Saharan and Middle Eastern lineages in Indian Muslim populations. *European Journal of Human Genetics*, 18(3), pp. 354–363.
Ebisch, I. et al., 2006. Does folic acid and zinc sulphate intervention affect endocrine parameters and sperm characteristics in men? *International journal of andrology*, 29(2), pp. 339–345.
Edwards, E.A. & Duntley, S.Q., 1939. The pigments and color of living human skin. *American Journal of Anatomy*, 65(1), pp. 1–33.
Enard, D., Messer, P.W. & Petrov, D.A., 2014. Genome-wide signals of positive selection in human evolution. *Genome research*.
Enattah, N.S. et al., 2007. Evidence of Still-Ongoing Convergence Evolution of the Lactase Persistence T< sub>-13910 Alleles in Humans. *The american journal of human genetics*, 81(3), pp. 615–625.

Enattah, N.S. et al., 2002. Identification of a variant associated with adult-type hypolactasia. *Nature genetics*, 30(2), pp. 233–237.
Enattah, N.S. et al., 2008. Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *The American Journal of Human Genetics*, 82(1), pp. 57–72.
Endicott, P. et al., 2003. The genetic origins of the Andaman Islanders. *The American Journal of Human Genetics*, 72(1), pp. 178–184.
Fanny, M.-P. et al., 2013. SLC24A5 mutations are associated with non-syndromic oculocutaneous albinism. *Journal of Investigative Dermatology*.
Fernandes, V. et al., 2012. The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa. *The American Journal of Human Genetics*, 90(2), pp. 347–355.
Field, J.S., Petraglia, M.D. & 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(1), pp. 88–108.
Fitzpatrick, T.B., 1988. The validity and practicality of sun-reactive skin types I through VI. *Archives of dermatology*, 124(6), pp. 869–871.
Flatz, G. & Rotthauwe, H., 1973. Lactose nutrition and natural selection. *The Lancet*, 302(7820), pp. 76–77.
Fleming, A., 2001. The role of folate in the prevention of neural tube defects: human and animal studies. *Nutrition reviews*, 59(8), pp. S13–S23.
Foll, M. et al., 2014. Hierarchical Bayesian model of population structure reveals convergent adaptation to high altitude in human populations. *arXiv preprint arXiv: 1402.4348*.
Forni, D. et al., 2013. A 175 million year history of T cell regulatory molecules reveals widespread selection, with adaptive evolution of disease alleles. *Immunity*, 38(6), pp. 1129–1141.
Fraser, H.B., 2013. Gene expression drives local adaptation in humans. *Genome research*, 23(7), pp. 1089–1096.
Frazer, K.A. et al., 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164), pp. 851–861.
Frazer, K.A. et al., 2009. Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics*, 10(4), pp. 241–251.
Frichot, E. et al., 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular biology and evolution*, 30(7), pp. 1687–1699.
Fuller, D.Q., 2006. Agricultural origins and frontiers in South Asia: a working synthesis. *Journal of World Prehistory*, 20(1), pp. 1–86.
Fumagalli, M. et al., 2009. Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. *The Journal of experimental medicine*, 206(6), pp. 1395–1408.
Fumagalli, M. et al., 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS genetics*, 7(11), p.e1002355.
Fumagalli, M. & Sironi, M., 2014. Human genome variability, natural selection and infectious diseases. *Current opinion in immunology*, 30, pp. 9–16.
Fu, W. & Akey, J.M., 2013. Selection and adaptation in the human genome. *Annual review of genomics and human genetics*, 14, pp. 467–489.

Gargiulo, A. et al., 2011. Molecular and clinical characterization of albinism in a large cohort of Italian patients. *Investigative ophthalmology & visual science*, 52(3), pp. 1281–1289.
Garn, S.M., Selby, S. & Crawford, M.R., 1956. Skin reflectance studies in children and adults. *American journal of physical anthropology*, 14(1), pp. 101–117.
Gerbault, P. et al., 2011. Evolution of lactase persistence: an example of human niche construction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1566), pp. 863–877.
Gerbault, P., 2014. The Onset of Lactase Persistence in Europe. *Human heredity*, 76(3-4), pp. 154–161.
Gettings, K.B. et al., 2014. A 50-SNP assay for biogeographic ancestry and phenotype prediction in the US population. *Forensic Science International: Genetics*, 8(1), pp. 101–108.
Ghaffar, A., Reddy, K.S. & Singhi, M., 2004. Burden of non-communicable diseases in South Asia. *BMJ*, 328(7443), pp. 807–810.
Ghosh, B., 2005. The Indian genome variation database (IGVdb): a project overview. *Human genetics*, 118(1), pp. 1–11.
Giardina, E. et al., 2008. Haplotypes in SLC24A5 gene as ancestry informative markers in different populations. *Current Genomics*, 9(2), p.110.
Ginger, R.S. et al., 2008. SLC24A5 encodes a trans-Golgi network protein with potassium-dependent sodium-calcium exchange activity that regulates human epidermal melanogenesis. *Journal of Biological Chemistry*, 283(9), pp. 5486–5495.
Glazko, G.V. & Nei, M., 2003. Estimation of divergence times for major lineages of primate species. *Molecular biology and evolution*, 20(3), pp. 424–434.
Gokhman, D. et al., 2014. Reconstructing the DNA Methylation Maps of the Neandertal and the Denisovan. *Science*, 344(6183), pp. 523–527.
Granka, J.M. et al., 2012. Limited evidence for classic selective sweeps in African populations. *Genetics*, 192(3), pp. 1049–1064.
Grant, W.B. & Holick, M.F., 2005. Benefits and requirements of vitamin D for optimal health: a review. *Altern Med Rev*, 10(2), pp. 94–111.
Green, R.E. et al., 2010. A draft sequence of the Neandertal genome. *science*, 328(5979), pp. 710–722.
Grigson, C., 1991. An African origin for African cattle? — some archaeological evidence. *African Archaeological Review*, 9(1), pp. 119–144.
Grønskov, K. et al., 2009. Birth prevalence and mutation spectrum in Danish patients with autosomal recessive albinism. *Investigative ophthalmology & visual science*, 50(3), pp. 1058–1064.
Grønskov, K., Ek, J. & Brondum-Nielsen, K., 2007. Oculocutaneous albinism. *Orphanet J Rare Dis*, 2, p.43.
Grossman, S.R. et al., 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*, 327(5967), pp. 883–886.
Grossman, S.R. et al., 2013. Identifying recent adaptations in large-scale genomic data. *Cell*, 152(4), pp. 703–713.
Günther, I. & Harttgen, K., 2013. *Desired Fertility and Children Born across Time and Space*, Courant Research Centre: Poverty, Equity and Growth-Discussion Papers.
Gupta, D. et al., 2007. Lactose intolerance in patients with irritable bowel syndrome from northern India: a case–control study. *Journal of gastroenterology and hepatology*, 22(12), pp. 2261–2265.

Hancock, A.M., Alkorta-Aranburu, G., et al., 2010a. Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1552), pp. 2459–2468.
Hancock, A.M., Witonsky, D.B., et al., 2010b. Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proceedings of the National Academy of Sciences*, 107(Supplement 2), pp. 8924–8930.
Harris, E.E. & Meyer, D., 2006. The molecular signature of selection underlying human adaptations. *American journal of physical anthropology*, 131(S43), pp. 89–130.
Harrison, G.A., 1957. The measurement and inheritance of skin colour in man. *The Eugenics review*, 49(2), p.73.
Harrison, G.A. & Owen, J., 1964. Studies on the inheritance of human skin colour. *Annals of human genetics*, 28(1-3), pp. 27–37.
Hart, K.L. et al., 2013. Improved eye-and skin-color prediction based on 8 SNPs. *Croatian medical journal*, 54(3), pp. 248–256.
Harvey, R. & Lord, J.M., 1978. Skin colour of the Ainu of Hidaka, Hokkaido, northern Japan. *Annals of human biology*, 5(5), pp. 459–467.
Hawks, J. et al., 2007. Recent acceleration of human adaptive evolution. *Proceedings of the National Academy of Sciences*, 104(52), pp. 20753–20758.
Hawley, J.S., 1979. Krishna's cosmic victories. *Journal of the American Academy of Religion*, 47(2), pp. 201–221.
Hawley, J.S., 1981. Scenes from the Childhood of Kṛṣṇa on the Kailāsanātha Temple, Ellora. *Archives of Asian Art*, pp. 74–90.
Henn, B.M. et al., 2010. Fine-scale population structure and the era of next-generation sequencing. *Human molecular genetics*, 19(R2), pp. R221–R226.
Henn, B.M. et al., 2012. Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS genetics*, 8(1), p.e1002397.
Heyer, E. et al., 2011. Lactase persistence in Central Asia: phenotype, genotype, and evolution. *Human biology*, 83(3), pp. 379–392.
Hinds, D.A. et al., 2005. Whole-genome patterns of common DNA variation in three human populations. *Science*, 307(5712), p.1072.
Holick, M.F., 2005. The vitamin D epidemic and its health consequences. *The Journal of nutrition*, 135(11), p.2739S–2748S.
Holick, M.F., 2003. Vitamin D: A millenium perspective. *Journal of cellular biochemistry*, 88(2), pp. 296–307.
Holick, M.F., 2007. Vitamin D deficiency. *New England Journal of Medicine*, 357(3), pp. 266–281.
Holick, M.F., 2004. Vitamin D: importance in the prevention of cancers, type 1 diabetes, heart disease, and osteoporosis. *The American journal of clinical nutrition*, 79(3), pp. 362–371.
Holick, M.F. & Chen, T.C., 2008. Vitamin D deficiency: a worldwide problem with health consequences. *The American journal of clinical nutrition*, 87(4), p.1080S–1086S.
Holick, M.F. & Jenkins, M., 2005. *The UV advantage*, ibooks.
Hollox, E.J. et al., 2001. Lactase haplotype diversity in the Old World. *The American Journal of Human Genetics*, 68(1), pp. 160–172.
Huerta-Sánchez, E. et al., 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*.
HUGO Pan-Asian SNP Consortium, 2009. Mapping human genetic diversity in Asia. *Science*, 326(5959), pp. 1541–1545.

Imtiaz, F. et al., 2007. The T/G− 13915 variant upstream of the lactase gene (LCT) is the founder allele of lactase persistence in an urban Saudi population. *Journal of medical genetics*, 44(10), pp. e89–e89.

Ingram, C.J., Elamin, M.F., et al., 2007. A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Human genetics*, 120(6), pp. 779–788.

Ingram, C.J., Mulcare, C.A., et al., 2009. Lactose digestion and the evolutionary genetics of lactase persistence. *Human genetics*, 124(6), pp. 579–591.

Ingram, C.J., Liebert, A. & Swallow, D.M., 2007. Population genetics of lactase persistence and lactose intolerance. *eLS*.

International HapMap Consortium, 2005. A haplotype map of the human genome. *Nature*, 437(7063), pp. 1299–1320.

Isaac, B.H., 2006. *The invention of racism in classical antiquity*, Princeton University Press.

Itan, Y. et al., 2010. A worldwide correlation of lactase persistence phenotype and genotypes. *BMC evolutionary biology*, 10(1), p.36.

Ito, S. & Wakamatsu, K., 2003. Quantitative analysis of eumelanin and pheomelanin in humans, mice, and other animals: a comparative review. *Pigment cell research*, 16(5), pp. 523–531.

Jablonski, D., Roy, K. & Valentine, J.W., 2006. Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science*, 314(5796), pp. 102–106.

Jablonski, N., 1999. A possible link between neural tube defects and ultraviolet light exposure. *Medical hypotheses*, 52(6), pp. 581–582.

Jablonski, N.G., 2004. The evolution of human skin and skin color. *Annual Review of Anthropology*, 33, pp. 585–623.

Jablonski, N.G. & Chaplin, G., 2010. Human skin pigmentation as an adaptation to UV radiation. *Proceedings of the National Academy of Sciences*, 107(Supplement 2), pp. 8962–8968.

Jablonski, N.G. & Chaplin, G., 2000a. The evolution of human skin coloration. *Journal of Human Evolution*, 39(1), pp. 57–106.

Jacobs, L.C. et al., 2013. Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. *Human genetics*, 132(2), pp. 147–158.

James, H.V.A. et al., 2005. Modern human origins and the evolution of behavior in the Later Pleistocene record of South Asia. Commentaries. Author's reply. *Current Anthropology*, 46, pp. S3–S27.

Järvelä, I., 2005. Molecular diagnosis of adult-type hypolactasia (lactase non-persistence). *Scandinavian journal of clinical & laboratory investigation*, 65(7), pp. 535–540.

Jarvis, J.P. et al., 2012. Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. *PLoS genetics*, 8(4), p.e1002641.

Jeong, C. et al., 2014. Admixture facilitates genetic adaptations to high altitude in Tibet. *Nature communications*, 5.

Jia, J. et al., 2014. Developing a novel panel of genome-wide ancestry informative markers for bio-geographical ancestry estimates. *Forensic Science International: Genetics*, 8(1), pp. 187–194.

Jorde, L., 2000. Linkage disequilibrium and the search for complex disease genes. *Genome research*, 10(10), pp. 1435–1444.

Kamberov, Y.G. et al., 2013. Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell*, 152(4), pp. 691–702.
Karafet, T.M. et al., 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome research*, 18(5), pp. 830–838.
Karlsson, E.K., Kwiatkowski, D.P. & Sabeti, P.C., 2014. Natural selection and infectious disease in human populations. *Nature Reviews Genetics*, 15(6), pp. 379–393.
Karve, I., 1968. *Kinship organization in India*, Asia Publishing House Bombay.
Kayser, M. et al., 2006. Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Molecular biology and evolution*, 23(11), pp. 2234–2244.
Kayser, M. et al., 2003. Y chromosome STR haplotypes and the genetic structure of US populations of African, European, and Hispanic ancestry. *Genome research*, 13(4), pp. 624–634.
Kayser, M. & de Knijff, P., 2011. Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Reviews Genetics*, 12(3), pp. 179–192.
Keller, A. et al., 2012. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature communications*, 3, p.698.
Kennedy, K.A. & Deraniyagala, S.U., 1989. Fossil remains of 28,000-year-old hominids from Sri Lanka. *Current Anthropology*, pp. 394–399.
Kennedy, K.A. & Langstroth, E., 2011. Recent Recovery of Unpublished Field Notes of Theodore D. McCown's Paleoanthropological Explorations in the Narmada River System, India, 1964–1965. *Asian Perspectives*, 50(1), pp. 132–143.
Kerckhove, E.V.D. et al., 2003. Reproducibility of repeated measurements on post-burn scars with Dermascan C. *Skin research and technology*, 9(1), pp. 81–84.
Keurentjes, J.J., Koornneef, M. & Vreugdenhil, D., 2008. Quantitative genetics in the age of omics. *Current opinion in plant biology*, 11(2), pp. 123–128.
Kierstein, G. et al., 2004. Analysis of mitochondrial D-loop region casts new light on domestic water buffalo (< i> Bubalus bubalis) phylogeny. *Molecular Phylogenetics and Evolution*, 30(2), pp. 308–324.
Kivisild, T. et al., 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *The American Journal of Human Genetics*, 72(2), pp. 313–332.
Kollias, N. et al., 1991. New trends in photobiology: Photoprotection by melanin. *Journal of Photochemistry and Photobiology B: Biology*, 9(2), pp. 135–160.
Kooner, J.S. et al., 2011. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nature genetics*, 43(10), pp. 984–989.
Kreitman, M., 2000. Methods to detect selection in populations with applications to the human. *Annual review of genomics and human genetics*, 1(1), pp. 539–559.
Krishnamoorthy, S. & Audinarayana, N., 2001. Trends in consanguinity in South India. *Journal of biosocial science*, 33(02), pp. 185–197.
Kruse, T. et al., 1988. The human lactase-phlorizin hydrolase gene is located on chromosome 2. *FEBS letters*, 240(1), pp. 123–126.
Kuijt, I. & Goring-Morris, N., 2002. Foraging, farming, and social complexity in the Pre-Pottery Neolithic of the southern Levant: a review and synthesis. *Journal of World Prehistory*, 16(4), pp. 361–440.

Kumar, S. et al., 2007. Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo. *Animal Genetics*, 38(3), pp. 227–232.

Kumar, V. et al., 2007. Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evolutionary Biology*, 7(1), p.47.

Kuokkanen, M. et al., 2003. Transcriptional regulation of the lactase-phlorizin hydrolase gene by polymorphisms associated with adult-type hypolactasia. *Gut*, 52(5), pp. 647–652.

Lacan, M. et al., 2011. Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proceedings of the National Academy of Sciences*, 108(45), pp. 18255–18259.

Lachance, J. et al., 2012. Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell*, 150(3), pp. 457–469.

Lahr, M.M. & Foley, R., 1994. Multiple dispersals and modern human origins. *Evolutionary Anthropology: Issues, News, and Reviews*, 3(2), pp. 48–60.

Laland, K.N., Odling-Smee, J. & Myles, S., 2010. How culture shaped the human genome: bringing genetics and the human sciences together. *Nature Reviews Genetics*, 11(2), pp. 137–148.

Lamason, R.L. et al., 2005. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science*, 310(5755), p.1782.

Lao, O. et al., 2007. Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Annals of Human Genetics*, 71(3), pp. 354–369.

Larson, G. & Burger, J., 2013. A population genetics view of animal domestication. *Trends in Genetics*, 29(4), pp. 197–205.

Lazaridis, I. et al., 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, 513(7518), pp. 409–413.

Lees, F.C. & Byard, P.J., 1978. Skin colorimetry in Belize. I. Conversion formulae. *American journal of physical anthropology*, 48(4), pp. 515–521.

Lee, V.A., Lorenz, K. & Singleton, A.D., 1979. The nutritional and physiological impact of milk in human nutrition. *Critical Reviews in Food Science & Nutrition*, 11(1), pp. 41–116.

Lehmuskallio, E., Hassi, J. & Kettunen, P., 2002. The skin in the cold. *International journal of circumpolar health*, 61(3).

Leonardi, M., 2013. Lactase persistence and milk consumption in Europe: an inter-disciplinary approach involving genetics and archaeology. *Documenta Praehisto-rica*, 40, pp. 84–96.

Lewinsky, R.H. et al., 2005. T− 13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. *Human molecular genetics*, 14(24), pp. 3945–3953.

Lewontin, R.C., 1995. The apportionment of human diversity. In *Evolutionary biology*. Springer, pp. 381–398.

Li, H., 2011. A new test for detecting recent positive selection that is free from the confounding impacts of demography. *Molecular biology and evolution*, 28(1), pp. 365–375.

Li, J. et al., 2012. Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Molecular ecology*, 21(1), pp. 28–44.

Li, J.Z. et al., 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *science*, 319(5866), pp. 1100–1104.

Li, M.J. et al., 2013. dbPSHP: a database of recent positive selection across human populations. *Nucleic acids research*, p.gkt1052.
Liu, F., Wen, B. & Kayser, M., 2013. Colorful DNA polymorphisms in humans. In Seminars in cell & developmental biology. Elsevier, pp. 562–575.
Loomis, W.F., 1967. Skin-pigment regulation of vitamin-D biosynthesis in man. *Science*, 157(3788), p.501.
Lopez, A.D. et al., 2006. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *The Lancet*, 367(9524), pp. 1747–1757.
Louicharoen, C. et al., 2009. Positively selected G6PD-Mahidol mutation reduces Plasmodium vivax density in Southeast Asians. *Science*, 326(5959), pp. 1546–1549.
Luca, F., Perry, G. & Di Rienzo, A., 2010. Evolutionary adaptations to dietary changes. *Annual review of nutrition*, 30, pp. 291–314.
Von Luschan, F., 1927. *Völker, rassen, sprachen: anthropologische betrachtungen*, Deutsche Buch-Gemeinschaft.
Lynch, M. & Walsh, B., 1998. Genetics and analysis of quantitative traits.
Macaulay, V. et al., 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science*, 308(5724), pp. 1034–1036.
Madrigal, L. & Kelly, W., 2006. Human skin-color sexual dimorphism: A test of the sexual selection hypothesis. *American journal of physical anthropology*, 132(3), pp. 470–482.
Magee, D.A., MacHugh, D.E. & Edwards, C.J., 2014. Interrogation of modern and ancient genomes reveals the complex domestic history of cattle. *Animal Frontiers*, 4(3), pp. 7–22.
Malmström, H. et al., 2010. High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. *BMC evolutionary biology*, 10(1), p.89.
Mangano, V.D. & Modiano, D., 2014. An evolutionary perspective of how infection drives human genome diversity: the case of malaria. *Current opinion in immunology*, 30, pp. 39–47.
Mantei, N. et al., 1988. Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme. *The EMBO journal*, 7(9), p.2705.
McCRACKEN, R.D., 1971. Origins and implications of the distribution of adult lactase deficiency in human populations. *Journal of Tropical Pediatrics*, 17(1), pp. 7–10.
McElreavey, K. & Quintana-Murci, L., 2005. A population genetics perspective of the Indus Valley through uniparentally-inherited markers. *Annals of human biology*, 32(2), pp. 154–162.
McEvoy, B., Beleza, S. & Shriver, M.D., 2006. The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Human molecular genetics*, 15(suppl 2), p.R176.
Meadow, R.H., 1993. Animal domestication in the Middle East: a revised view from the eastern margin. *Harappan civilization*, pp. 295–320.
Mehta, N.N., 2011. A genome-wide association study in Europeans and South Asians identifies 5 new loci for coronary artery disease. *Circulation: Cardiovascular Genetics*, 4(4), pp. 465–466.
Mellars, P. et al., 2013. Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. *Proceedings of the National Academy of Sciences*, 110(26), pp. 10699–10704.

Mellars, P., 2006. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science*, 313(5788), pp. 796–800.
Mendez, F.L., Watkins, J.C. & Hammer, M.F., 2012. A Haplotype at< i> STAT2 Introgressed from Neanderthals and Serves as a Candidate of Positive Selection in Papua New Guinea. *The American Journal of Human Genetics*, 91(2), pp. 265–274.
Mendez, F.L., Watkins, J.C. & Hammer, M.F., 2013. Neandertal origin of genetic variation at the cluster of OAS immunity genes. *Molecular biology and evolution*, 30(4), pp. 798–801.
Metspalu, M. et al., 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC genetics*, 5(1), p.26.
Metspalu, M. et al., 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(6), pp. 731–744.
Meyer, M. et al., 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science*, 338(6104), pp. 222–226.
Miller, C.T. et al., 2007. cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell*, 131(6), pp. 1179–1189.
Mitchell, J. & Collinson, P., 1744. An Essay upon the Causes of the Different Colours of People in Different Climates; By John Mitchell, MD Communicated to the Royal Society by Mr. Peter Collinson, FRS. *Philosophical Transactions*, 43(472–477), pp. 102–150.
Mondal, M. et al., 2012. Molecular basis of albinism in India: Evaluation of seven potential candidate genes and some new findings. *Gene*, 511(2), pp. 470–474.
Montoliu, L. et al., 2014. Increasing the complexity: new genes and new types of albinism. *Pigment cell & melanoma research*, 27(1), pp. 11–18.
Moorjani, P. et al., 2013. Genetic evidence for recent population mixture in India. *The American Journal of Human Genetics*, 93(3), pp. 422–438.
Mulcare, C.A. et al., 2004. The T Allele of a Single-Nucleotide Polymorphism 13.9 kb Upstream of the Lactase Gene (< i> LCT)(< i> C− 13.9 kbT) Does Not Predict or Cause the Lactase-Persistence Phenotype in Africans. *The American Journal of Human Genetics*, 74(6), pp. 1102–1110.
Myles, S. et al., 2005. Genetic evidence in support of a shared Eurasian-North African dairying origin. *Human genetics*, 117(1), pp. 34–42.
Myles, S. et al., 2008. Identification and Analysis of Genomic Regions with Large Between-Population Differentiation in Humans. *Annals of human genetics*, 72(1), pp. 99–110.
Naegele, C.J., 2008. History and influence of law code of Manu.
Naik, S. et al., 2013. Daily milk intake improves vitamin B-12 status in young vegetarian Indians: an intervention trial. *Nutrition journal*, 12(1), p.136.
Naik, S., 1978. Origin and domestication of Zebu cattle (< i> Bos indicus). *Journal of Human Evolution*, 7(1), pp. 23–30.
Narang, A. et al., 2010. IGVBrowser–a genomic variation resource from diverse Indian populations. *Database*, 2010, p.baq022.
Narang, A. et al., 2011. Recent admixture in an Indian population of African ancestry. *The American Journal of Human Genetics*, 89(1), pp. 111–120.
Neel, J.V., 1962. Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? *American journal of human genetics*, 14(4), p.353.
Neel, J.V., 1999. The “Thrifty Genotype” in 19981. *Nutrition reviews*, 57(5), pp. 2–9.

Nei, M., 1985. Human evolution at the molecular level. *Population genetics and molecular evolution*, 1, pp. 41–64.
Nessvi, S. et al., 2011. Association of 25-Hydroxyvitamin D3 Levels in Adult New Zealanders with Ethnicity, Skin Color and Self-Reported Skin Sensitivity to Sun Exposure. *Photochemistry and photobiology*, 87(5), pp. 1173–1178.
Nielsen, R., 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.*, 39, pp. 197–218.
Norton, H.L. et al., 2007. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Molecular biology and evolution*, 24(3), p.710.
Novembre, J. & Di Rienzo, A., 2009. Spatial patterns of variation due to natural selection in humans. *Nature Reviews Genetics*, 10(11), pp. 745–755.
Okoro, A., 1975. Albinism in Nigeria1. *British Journal of Dermatology*, 92(5), pp. 485–492.
Olalde, I. et al., 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature*, 507(7491), pp. 225–228.
Olds, L.C. & Sibley, E., 2003. Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a cis regulatory element. *Human molecular genetics*, 12(18), pp. 2333–2340.
Olson, M.V. & Varki, A., 2003. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nature Reviews Genetics*, 4(1), pp. 20–28.
Oppenheimer, S., 2012. A single southern exit of modern humans from Africa: Before or after Toba? *Quaternary International*, 258, pp. 88–99.
Ortonne, J., 2002. Photoprotective properties of skin melanin. *British Journal of Dermatology*, 146(s61), pp. 7–10.
Paddayya, K., 2001. The Acheulian culture project of the Hunsgi and Baichbal Valleys, peninsular India. *Human Roots: Africa and Asia in the Middle Pleistocene*, pp. 235–58.
Pala, M. et al., 2012. Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *The American Journal of Human Genetics*, 90(5), pp. 915–924.
Palanichamy, M.G. et al., 2010. Mitochondrial haplogroup N1a phylogeography, with implication to the origin of European farmers. *BMC evolutionary biology*, 10(1), p.304.
Palanichamy, M. gounder et al., 2004. Phylogeny of Mitochondrial DNA Macrohaplo-group N in India, Based on Complete Sequencing: Implications for the Peopling of South Asia. *The American Journal of Human Genetics*, 75(6), pp. 966–978.
Paradies, Y., Montoya, M.J. & Fullerton, S.M., 2007. Racialized genetics and the study of complex diseases: the thrifty genotype revisited. *Perspectives in biology and medicine*, 50(2), pp. 203–227.
Parra, E.J., 2007. Human pigmentation variation: evolution, genetic basis, and implications for public health. *American journal of physical anthropology*, 134(S45), pp. 85–105.
Patel, A.K. & Meadow, R., 1998. The exploitation of wild and domestic water buffalo in prehistoric northwestern South Asia. *Archaeozoology of the Near East III*, pp. 180–199.
Paterson, S., 1998. Evidence for balancing selection at the major histocompatibility complex in a free-living ruminant. *Journal of Heredity*, 89(4), pp. 289–294.
Pathak, M.A., 2004. In memory of Thomas Bernhard Fitzpatrick. *Journal of Investigative Dermatology*, 122(2), pp. xx–xxi.

Peng, M.-S. et al., 2012. Lactase persistence may have an independent origin in Tibetan populations from Tibet, China. *Journal of human genetics*, 57(6), pp. 394–397.
Perry, G.H. et al., 2007. Diet and the evolution of human amylase gene copy number variation. *Nature genetics*, 39(10), pp. 1256–1260.
Peter, B.M., Huerta-Sanchez, E. & Nielsen, R., 2012. Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS genetics*, 8(10), p.e1003011.
Petraglia, M. et al., 2007. Middle Paleolithic assemblages from the Indian subcontinent before and after the Toba super-eruption. *Science*, 317(5834), pp. 114–116.
Petraglia, M.D. & Allchin, B., 2007. *The evolution and history of human populations in South Asia*, Springer.
Phillips, C. et al., 2014. Building a forensic ancestry panel from the ground up: The EUROFORGEN Global AIM-SNP set. *Forensic Science International: Genetics*, 11, pp. 13–25.
Phillips, C. et al., 2007. Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Science International: Genetics*, 1(3-4), pp. 273–280.
Pickrell, J.K. et al., 2009. Signals of recent positive selection in a worldwide sample of human populations. *Genome Research*, 19(5), pp. 826–837.
Piel, F.B. et al., 2010. Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nature communications*, 1, p.104.
Pinhasi, R., Fort, J. & Ammerman, A.J., 2005. Tracing the origin and spread of agriculture in Europe. *PLoS biology*, 3(12), p.e410.
Plantinga, T.S. et al., 2012. Low prevalence of lactase persistence in Neolithic South-West Europe. *European journal of human genetics*, 20(7), pp. 778–782.
Pneuman, A. et al., 2012. Verification of eye and skin color predictors in various populations. *Legal Medicine*, 14(2), pp. 78–83.
Post, P.W., 1975. Introduction: the study of pigmentation. *American Journal of Physical Anthropology*, 43(3), pp. 382–386.
Poulter, M. et al., 2003. The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. *Annals of human genetics*, 67(4), pp. 298–311.
Pritchard, J.K., Pickrell, J.K. & Coop, G., 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, 20(4), pp. R208–R215.
Prüfer, K. et al., 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505(7481), pp. 43–49.
Prugnolle, F. et al., 2005. Dispersal in a parasitic worm and its two hosts: consequence for local adaptation. *Evolution*, 59(2), pp. 296–303.
Quillen, E.E. & Shriver, M.D., 2011. Milestone 2. *Nature Milestones*, pp. E5–E7.
Abi-Rached, L. et al., 2011. The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science*, 334(6052), pp. 89–94.
Radha, V. & Mohan, V., 2007. Genetic predisposition to type 2 diabetes among Asian Indians. *The Indian journal of medical research*, 125(3), pp. 259–74.
Radwan, J. & Babik, W., 2012. The genomics of adaptation. *Proceedings of the Royal Society B: Biological Sciences*, 279(1749), pp. 5024–5028.
Rai, N. et al., 2012. The phylogeography of Y-chromosome haplogroup h1a1a-m82 reveals the likely Indian origin of the European Romani populations. *PloS one*, 7(11), p.e48477.

Raj, S.M. et al., 2013. A general linear model-based approach for inferring selection to climate. *BMC genetics*, 14(1), p.87.
Ranciaro, A. et al., 2014. Genetic origins of lactase persistence and the spread of pastoralism in Africa. *The American Journal of Human Genetics*, 94(4), pp. 496–510.
Rasinperä, H. et al., 2004. A genetic test which can be used to diagnose adult-type hypolactasia in children. *Gut*, 53(11), pp. 1571–1576.
Reddy, V. & Pershad, J., 1972. Lactase deficiency in Indians. *The American journal of clinical nutrition*, 25(1), pp. 114–119.
Reich, D. et al., 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *The American Journal of Human Genetics*, 89(4), pp. 516–528.
Reich, D. et al., 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468(7327), pp. 1053–1060.
Reich, D. et al., 2009. Reconstructing Indian population history. *Nature*, 461(7263), pp. 489–494.
Relethford, J.H., 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. *American Journal of Physical Anthropology*, 118(4), pp. 393–398.
Relethford, J.H., 1994. Craniometric variation among modern human populations. *American Journal of Physical Anthropology*, 95(1), pp. 53–62.
Relethford, J.H., 1997. Hemispheric difference in human skin color. *American journal of physical anthropology*, 104(4), pp. 449–457.
Relethford, J.H., 2000. Human skin color diversity is highest in sub-Saharan African populations. *Human biology*, pp. 773–780.
Relethford, J.H., 2009. Race and global patterns of phenotypic variation. *American journal of physical anthropology*, 139(1), pp. 16–22.
Relethford, J.H. et al., 1983. Social class, admixture, and skin color variation in Mexican-Americans and Anglo-Americans living in San Antonio, Texas. *American Journal of Physical Anthropology*, 61(1), pp. 97–102.
Renfrew, C., 2006. Inception of agriculture and rearing in the Middle East. *Comptes Rendus Palevol*, 5(1), pp. 395–404.
Richards, M. et al., 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *The American Journal of Human Genetics*, 67(5), pp. 1251–1276.
Richerson, P.J. & Boyd, R., 2008. *Not by genes alone: How culture transformed human evolution*, University of Chicago Press.
Di Rienzo, A. & Hudson, R.R., 2005. An evolutionary framework for common diseases: the ancestral-susceptibility model. *TRENDS in Genetics*, 21(11), pp. 596–601.
Roberts, D. & Kahlon, D., 1976. Environmental correlations of skin colour. *Annals of human biology*, 3(1), pp. 11–22.
Robins, A.H., 2005. *Biological perspectives on human pigmentation*, Cambridge University Press.
Robinson, D.W., Korisettar, R. & Koshy, J., 2008. Metanarratives and the (re) invention of the Neolithic A case study in rock-art from Birappa Rock Shelter and Hiregudda Hill, South-Central India. *Journal of Social Archaeology*, 8(3), pp. 355–379.
Robinson, M.R., Wray, N.R. & Visscher, P.M., 2014. Explaining additional genetic variation in complex traits. *Trends in Genetics*, 30(4), pp. 124–132.
Rockman, M.V., 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, 66(1), pp. 1–17.

Ronald, J. & Akey, J.M., 2005. Genome-wide scans for loci under selection in humans. *Human genomics*, 2(2), p.113.

Rosdahl, I. & Szabo, G., 1976. Ultrastructure and Biochemical Organization of the Pigmentary System. In Proceedings of the 9th International Pigment Cell Conference, Houston, Tex., January 13-17, 1975: Unique properties of melanocytes. S Karger Ag, p. 1.

Rubegni, P. et al., 1999. Quantitative Characterization and Study of the Relationship between Constitutive-Facultative Skin Color and Phototype in Caucasians. *Photochemistry and photobiology*, 70(3), pp. 303–307.

Ryman, N., Chakraborty, R. & Nei, M., 1983. Differences in the relative distribution of human gene diversity between electrophoretic and red and white cell antigen loci. *Human heredity*, 33(2), pp. 93–102.

Sääf, M. et al., 2011. Severe vitamin D deficiency in pregnant women of Somali origin living in Sweden. *Acta Paediatrica*, 100(4), p.612.

Sabbagh, A. et al., 2008. Worldwide distribution of NAT2 diversity: implications for NAT2 evolutionary history. *BMC genetics*, 9(1), p.21.

Sabbagh, A. & Darlu, P., 2005. Inferring haplotypes at the NAT2 locus: the computational approach. *BMC genetics*, 6(1), p.30.

Sabeti, P. et al., 2006. Positive natural selection in the human lineage. *science*, 312(5780), pp. 1614–1620.

Sabeti, P.C. et al., 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909), pp. 832–837.

Sabeti, P.C. et al., 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164), pp. 913–918.

Sahi, T., 1994. Genetics and epidemiology of adult-type hypolactasia. *Scandinavian Journal of Gastroenterology*, 29(S202), pp. 7–20.

Sahi, T., 1973. The inheritance of selective adult-type lactose malabsorption. *Scandinavian journal of gastroenterology. Supplement*, 30, pp. 1–73.

Sahi, T. & Launiala, K., 1977. More evidence for the recessive inheritance of selective adult type lactose malabsorption. *Gastroenterology*, 73(2), pp. 231–232.

Sahoo, S. et al., 2006. A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proceedings of the National Academy of Sciences of the United States of America*, 103(4), pp. 843–848.

Sanghera, D.K. et al., 2008. Impact of nine common type 2 diabetes risk polymorphisms in Asian Indian Sikhs: PPARG2 (Pro12Ala), IGF2BP2, TCF7L2 and FTO variants confer a significant risk. *BMC medical genetics*, 9(1), p.59.

Sankararaman, S. et al., 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, 507(7492), pp. 354–357.

Sankhyan, A., 1997. Fossil clavicle of a middle Pleistocene hominid from the central Narmada Valley, India. *Journal of human evolution*, 32(1), pp. 3–16.

Sankhyan, A., 2005. New fossils of early stone age man from central Narmada valley. *Current Science*, 88(5), pp. 704–707.

Sankhyan, A.R., 1997. A new human fossil find from the Central Narmada basin and its chronology. *Current science*, 73(12), pp. 1110–1111.

Sankhyan, A.R. et al., 2012. New Postcranial Hominin Fossils from the Central Narmada Valley, India. *Advances in Anthropology*, 2(03), p.125.

Saunders, M.A., 2004. Patterns of nucleotide variability within and around G6PD, a locus under positive natural selection in humans.

Saxena, R. et al., 2012. Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. *The American Journal of Human Genetics*, 90(3), pp. 410–425.

Scheinfeldt, L.B. & Tishkoff, S.A., 2013. Recent human adaptation: genomic approaches, interpretation and insights. *Nature Reviews Genetics*, 14(10), pp. 692–702.

Schnorr, S.L. et al., 2014. Gut microbiome of the Hadza hunter-gatherers. *Nature communications*, 5.

Scrimshaw, N.S. & Murray, E.B., 1988. The acceptability of milk and milk products in populations with a high prevalence of lactose intolerance. *The American journal of clinical nutrition*, 48(4), pp. 1079–1159.

Sengupta, M. et al., 2010. Comprehensive analysis of the molecular basis of oculocutaneous albinism in Indian patients lacking a mutation in the tyrosinase gene. *British Journal of Dermatology*, 163(3), pp. 487–494.

Sengupta, M., 2009. Genetic and Molecular Basis of Albinism in Indian Patients.

Sengupta, S. et al., 2006. Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *The American Journal of Human Genetics*, 78(2), pp. 202–221.

Sequencing, T.C. & Analysis Consortium, 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437(7055), pp. 69–87.

Serre, D. & Pääbo, S., 2004. Evidence for gradients of human genetic diversity within and among continents. *Genome research*, 14(9), pp. 1679–1685.

Shah, A.M. et al., 2011. Indian siddis: African descendants with Indian admixture. *The American Journal of Human Genetics*, 89(1), pp. 154–161.

Sharma, V.P., Köhler-Rollefson, I. & Morton, J., 2003. Pastoralism in India: a scoping study. *Indian Institute of Management and League of Pastoral Peoples, Ahmedabad, India and Ober-Ramstadt, Germany*, 236.

Shriver, M.D. et al., 2003. Skin pigmentation, biogeographical ancestry and admixture mapping. *Human Genetics*, 112(4), pp. 387–399.

Sikora, M. et al., 2014. Population Genomic Analysis of Ancient and Modern Genomes Yields New Insights into the Genetic Ancestry of the Tyrolean Iceman and the Genetic Structure of Europe. *PLoS genetics*, 10(5), p.e1004353.

Simonsen, K.L., Churchill, G.A. & Aquadro, C.F., 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*, 141(1), pp. 413–429.

Simoons, F.J., 1970a. Primary adult lactose intolerance and the milking habit: a problem in biologic and cultural interrelations. *The American journal of digestive diseases*, 15(8), pp. 695–710.

Simoons, F.J., 1970b. The traditional limits of milking and milk use in southern Asia. *Anthropos*, pp. 547–593.

Sims, L.M. & Ballantyne, J., 2008. The golden gene (SLC24A5) differentiates US subpopulations within the ethnically admixed Y-SNP haplogroups. *Legal Medicine*, 10(2), pp. 72–77.

Singh, K. & Pundir, R., 2000. *Co-operatives and Rural Development in India*, Institute of Rural Management.

Skoglund, P. et al., 2014. Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science*, 344(6185), pp. 747–750.

Smith, S.S., 1965. *An Essay on the Causes of the Variety of Complexion and Figure in the Human Species*, Belknap Press of Harvard University Press.

Soejima, M. & Koda, Y., 2007. Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC45A2. *International Journal of Legal Medicine*, 121(1), pp. 36–39.

Sonakia, A. & Kennedy, K.A., 1985. Skull cap of an early man from the Narmada valley alluvium (Pleistocene) of central India. *American Anthropologist*, 87(3), pp. 612–616.

Spichenok, O. et al., 2011. Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Science International: Genetics*, 5(5), pp. 472–478.

Stamatas, G.N. et al., 2004. Non-Invasive Measurements of Skin Pigmentation In Situ. *Pigment Cell Research*, 17(6), pp. 618–626.

Steiper, M.E. & Young, N.M., 2006. Primate molecular divergence dates. *Molecular phylogenetics and evolution*, 41(2), pp. 384–394.

Stewart, J.R. & Stringer, C.B., 2012. Human evolution out of Africa: the role of refugia and climate change. *Science*, 335(6074), pp. 1317–1321.

Stokowski, R.P. et al., 2007. A genomewide association study of skin pigmentation in a South Asian population. *The American Journal of Human Genetics*, 81(6), pp. 1119–1132.

Stoneking, M. & Krause, J., 2011. Learning about human population history from ancient and modern genomes. *Nature Reviews Genetics*, 12(9), pp. 603–614.

Stringer, C., 2014. Why we are not all multiregionalists now. *Trends in ecology & evolution*, 29(5), pp. 248–251.

Sturm, R.A., Box, N.F. & Ramsay, M., 1998. Human pigmentation genetics: the difference is only skin deep. *Bioessays*, 20(9), pp. 712–721.

Sturm, R.A. & Duffy, D.L., 2013. Human pigmentation: painting by numbers or ancestry? *Pigment Cell & Melanoma Research*, 26(5), pp. 605–606.

Sun, C. et al., 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(6), pp. 966–978.

Sverrisdóttir, O.Ó. et al., 2014. Direct estimates of natural selection in Iberia indicate calcium absorption was not the only driver of lactase persistence in Europe. *Molecular biology and evolution*, p.msu049.

Swallow, D.M., 2003. Genetics of lactase persistence and lactose intolerance. *Annual review of genetics*, 37(1), pp. 197–219.

Swaminathan, N. et al., 1970. Disaccharidase levels in je junal biopsy specimens from American and South Indian control subjects and patients with tropical sprue. *Clinica Chimica Acta*, 30(3), pp. 707–712.

Swiatoniowski, A.K. et al., 2013. Comparing von Luschan skin color tiles and modern spectrophotometry for measuring human skin pigmentation. *American journal of physical anthropology*.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), pp. 585–595.

Takiwaki, H. et al., 2002. Skin reflectance-spectra and colour-value dependency on measuring-head aperture area in ordinary reflectance spectrophotometry and tristimulus colourimetry. *Skin Research and Technology*, 8(2), pp. 94–97.

Tamang, R., Singh, L. & Thangaraj, K., 2012. Complex genetic origin of Indian populations and its implications. *Journal of biosciences*, 37(5), pp. 911–919.

Tamang, R. & Thangaraj, K., 2012. Genomic view on the peopling of India. *Investigative genetics*, 3(1), p.20.

Tandon, R.K. et al., 1981. Lactose intolerance in North and South Indians. *The American journal of clinical nutrition*, 34(5), pp. 943–946.
Tang, K., Thornton, K.R. & Stoneking, M., 2007. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS biology*, 5(7), p.e171.
Tasa, G.L., Murray, C.J. & Boughton, J.M., 1985. Reflectometers Reports on Human Pigmentation. *Current anthropology*, 26(4), pp. 511–512.
Taylor, S. et al., 2006. Noninvasive techniques for the evaluation of skin color. *Journal of the American Academy of Dermatology*, 54(5), pp. S282–S290.
Teshima, K.M. & Przeworski, M., 2006. Directional positive selection on an allele of arbitrary dominance. *Genetics*, 172(1), pp. 713–718.
Thangaraj, K. et al., 2006. In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup'M'in India. *BMC genomics*, 7(1), p.151.
Thangaraj, K. et al., 2008. Maternal footprints of southeast Asians in North India. *Human heredity*, 66(1), pp. 1–9.
Thangaraj, K. et al., 2005. Reconstructing the origin of Andaman Islanders. *Science*, 308(5724), pp. 996–996.
Thody, A.J. et al., 1991. Pheomelanin as well as eumelanin is present in human epidermis. *Journal of Investigative Dermatology*, 97(2), pp. 340–344.
Thong, H. et al., 2003. The patterns of melanosome distribution in keratinocytes of human skin as one determining factor of skin colour. *British Journal of Dermatology*, 149(3), pp. 498–505.
Tishkoff, S.A. et al., 2006. Convergent adaptation of human lactase persistence in Africa and Europe. *Nature genetics*, 39(1), pp. 31–40.
Tishkoff, S.A. et al., 2001. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science*, 293(5529), pp. 455–462.
Tishkoff, S.A. et al., 2009. The genetic structure and history of Africans and African Americans. *Science*, 324(5930), pp. 1035–1044.
Tishkoff, S.A. & Kidd, K.K., 2004. Implications of biogeography of human populations for'race'and medicine. *Nature genetics*, 36, pp. S21–S27.
Tishkoff, S.A. & Verrelli, B.C., 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annual review of genomics and human genetics*, 4(1), pp. 293–340.
Tobias, P., 1978. Reflections on Anatomy and Physical Anthropology. *South African Medical Journal*, 53, pp. 1066–1071.
Toda, K. et al., 1972. Alteration of racial differences in melanosome distribution in human epidermis after exposure to ultraviolet light. *Nature*, 236(66), pp. 143–145.
Tomita, Y. & Suzuki, T., 2004. Genetics of pigmentary disorders. In American Journal of Medical Genetics Part C: Seminars in Medical Genetics. Wiley Online Library, pp. 75–81.
Torniainen, S. et al., 2009. Screening of variants for lactase persistence/non-persistence in populations from South Africa and Ghana. *BMC genetics*, 10(1), p.31.
Tresset, A. & Vigne, J.-D., 2011. Last hunter-gatherers and first farmers of Europe. *Comptes rendus biologies*, 334(3), pp. 182–189.
Troelsen, J.T. et al., 2003. An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology*, 125(6), pp. 1686–1694.
Tsetskhladze, Z.R. et al., 2012. Functional Assessment of Human Coding Mutations Affecting Skin Pigmentation Using Zebrafish. *PloS one*, 7(10), p.e47398.

Underhill, P.A. et al., 2014. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *European Journal of Human Genetics*.
Underhill, P.A. & Kivisild, T., 2007. Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu. Rev. Genet.*, 41, pp. 539–564.
Valenzuela, R.K. et al., 2010. Predicting Phenotype from Genotype: Normal Pigmentation*. *Journal of forensic sciences*, 55(2), pp. 315–322.
Varki, A. & Altheide, T.K., 2005. Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Research*, 15(12), pp. 1746–1758.
Veeramah, K.R. & Hammer, M.F., 2014. The impact of whole-genome sequencing on the reconstruction of human population history. *Nature Reviews Genetics*, 15(3), pp. 149–162.
Vernot, B. & Akey, J.M., 2014. Resurrecting surviving neandertal lineages from modern human genomes. *Science*, 343(6174), pp. 1017–1021.
Vigne, J.-D., 2011. The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. *Comptes rendus biologies*, 334(3), pp. 171–181.
Visser, M., Kayser, M. & Palstra, R.-J., 2012. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome research*, 22(3), pp. 446–455.
Vogel, P. et al., 2008. Ocular Albinism and Hypopigmentation Defects in Slc24a5—I—Mice. *Veterinary Pathology Online*, 45(2), pp. 264–279.
Voight, B.F. et al., 2006. A map of recent positive selection in the human genome. *PLoS biology*, 4(3), p.e72.
Wang, H. et al., 2007. Comparative and evolutionary pharmacogenetics of ABCB1: complex signatures of positive selection on coding and regulatory regions. *Pharmacogenetics and genomics*, 17(8), pp. 667–678.
Wang, Y. et al., 1998. The genetically programmed down-regulation of lactase in children. *Gastroenterology*, 114(6), pp. 1230–1236.
Wang, Y. et al., 1995. The lactase persistence/non-persistence polymorphism is controlled by a cis-acting element. *Human molecular genetics*, 4(4), pp. 657–662.
Watkins, W. et al., 2008. Genetic variation in South Indian castes: evidence from Y-chromosome, mitochondrial, and autosomal polymorphisms. *BMC genetics*, 9(1), p.86.
Wei, A.-H. et al., 2013. Exome sequencing identifies SLC24A5 as a candidate gene for nonsyndromic oculocutaneous albinism. *Journal of Investigative Dermatology*, 133(7), pp. 1834–1840.
Weiner, J., 1951. 253. A Spectrophotometer for Measurement of Skin Colour. *Man*, pp. 152–153.
Westerhof, W., 1995. CIE colorimetry. *Handbook of non-invasive methods and the skin*, pp. 385–397.
Whiteman, D.C., Parsons, P.G. & Green, A.C., 1999. Determinants of melanocyte density in adult human skin. *Archives of dermatological research*, 291(9), pp. 511–516.
Wilde, S. et al., 2014. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proceedings of the National Academy of Sciences*, 111(13), pp. 4832–4837.
Williams, T.N. & Weatherall, D.J., 2012. World distribution, population genetics, and health burden of the hemoglobinopathies. *Cold Spring Harbor perspectives in medicine*, 2(9), p.a011692.

Witzel, M., 2005. Central Asian roots and acculturation in South Asia: linguistic and archaeological evidence from western central Asia, the Hindukush and northwestern south Asia for early Indo-Aryan language and religion. *Liguistics, archaeology and the human past*, pp. 87–211.

Wolfe, N.D., Dunavan, C.P. & Diamond, J., 2007. Origins of major human infectious diseases. *Nature*, 447(7142), pp. 279–283.

Wray, G.A., 2007. The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, 8(3), pp. 206–216.

Wu, W., 2012. Modern quantitative genetics: Dissecting complex polygenic systems into individual genetic factors. *Chinese Science Bulletin*, 57(21), pp. 2635–2636.

Xing, J. et al., 2009. Fine-scaled human genetic structure revealed by SNP microarrays. *Genome research*, 19(5), pp. 815–825.

Xing, J. et al., 2010. Genetic diversity in India and the inference of Eurasian population expansion. *Genome Biology*, 11(11), p.R113.

Yadav, S. et al., 2013. Detailed analysis of gene polymorphisms associated with ischemic stroke in South Asians. *PloS one*, 8(3), p.e57305.

Yi, X. et al., 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, 329(5987), pp. 75–78.

Young, A.R. & Sheehan, J.M., 2001. UV-induced pigmentation in human skin. *Comprehensive Series in Photosciences*, 3, pp. 357–375.

Young, J.H. et al., 2005. Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion. *PLoS genetics*, 1(6), p.e82.

Yunusbayev, B. et al., 2012. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Molecular biology and evolution*, 29(1), pp. 359–365.

Zhivotovsky, L.A. et al., 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *The American Journal of Human Genetics*, 74(1), pp. 50–61.

Zhou, D. et al., 2013. Whole-genome sequencing uncovers the genetic basis of chronic mountain sickness in Andean highlanders. *The American Journal of Human Genetics*, 93(3), pp. 452–462.

ACKNOWLEDGEMENTS

I feel that "getting graduate" is an academic degree for sure but the process to achieve it is like a small adventurous journey through the trail. You walk, you stop, you fall, you get up again, you see something, get very enthusiastic to know about it, you explore, you learn, you try and try again. So, before I step towards the completion of this journey of mine, I would like to acknowledge with all my heart, to all the people who helped me reach the finish line and I am extremely grateful to them for their support and encouragement.

First and foremost, I wish to thank my supervisors, Prof. Toomas Kivisild and Prof. Richard Villems who have been my guides for the journey. They nurtured me with their own experience and gave me the independence to choose the path I wanted. They never imposed their decisions on me, rather helped me so that I could evolve to handle the problems independently and try to solve them on my own and gave suggestions and directions wherever needed. I thank them for all their patience and support.

I would like to thank my seniors – Ene, Siiri and Maere, who were there to help me whenever needed. I am grateful to Mait for his cordial support and advice. He was also with me during one of the field trips to India. I would like to thank all my colleagues who gave me such a great company and created a warm and congenial atmosphere at the work. I never felt as an alien in the lab and also anytime during my stay here. They were always willing to help me with local things, to know their culture and at work, I always felt like a part of the great team. Lena, thank you for your unconditional support and motivation during my thesis writing and constant supply of cheese and bread at times when I was working off hours ☺. Jüri, you are like our Santa Claus (caring for everybody around selflessly) and pictures taken by you will be a real asset for me. Georgi, thanks for the useful discussions over the project. Monika, thank you for constantly cheering me up and I cherish the first glimpses of Estonian country-side with you. Ajai, Monika, Kristiina, Mari, Erika, Erwan, Anu, Eva-Liis, Hovo, Erik, Anne-Mai, Helle-Viivi, Lauri, Bayazit – thank you all for your support, academic discussions and those real happy moments, both over the work and outside (conference trips, sledging, picnics, dinners, Estonian nature and barbeques). My colleagues also helped me with the Estonian bits of the thesis and I am grateful to them for this. My sincere thanks to Ille, Tuuli and Trinu for their technical help, still remember running down with sequencing plates. I am grateful to our visiting fellows who made the environment yet more international – Lejla, Nina, Tena, Jelena, Dubravka, Tonka, Anna, Mannis, Federica, Irene, Nastia, Sardana, Ildus and Nataliya. I loved the moments spent with you all and laughters in between the work. I wish to render special note of thanks to my friends outside the lab - Aditi, Vijay Sir, Archana, Evelyn and Clotilde for supporting me and cheering me up.

I am also deeply grateful to my previous supervisors Dr K. Thangaraj and Dr. R. Ramani with whom I had my first introductions in the field and it was a great learning experience with them.

I am thankful to my collaborators and co-authors for their help in collection of samples, suggestions and scholarly inputs during manuscript preparation, in particular to Mircea (It was great working together!) and Märt Möls who guided me with the statistical analyses and was always patient to answer my questions. I would also like to thank Prof. Maris Laan for reviewing my thesis and for her useful comments and suggestions, which further helped to improve the dissertation.

I would like to thank my Indian friends who form a small niche here. They were like my extended family during these years.

Last but not the least, my sincerest gratitude to my parents, in-laws, brother (his family), for being a constant source of encouragement and strength, believing in me and helping me to hold it there and supporting me in all my pursuits. Heartful thanks to my husband Gyaneshwer, who has been my greatest supporter as well as my critic. I really appreciate all your patience, understanding and unconditional love during these years and especially during the last stages of writing. Thank you for your unwavering support and it is an absolute bliss to have you in my life. Hugs to my three year old tiny tot Vedant (Nuppu) – your smile means the world to me. Thank you for being so loving and co-operative. You travelled with us in one of our conference trips in your first year and your expression of "Dubrovnik" is still afresh on our minds ☺.

PUBLICATIONS

CURRICULUM VITAE

Name: Chandana Basu Mallick
Date of Birth: 25th July 1980
Address: Department of Evolutionary Biology,
Institute of Molecular and Cell Biology,
Riia 23B, Tartu, Estonia, 51010
Phone (Office): +372 737 5005
E-mail: cbm2577@gmail.com

Academic Background:

(April 2007 – till present)	PhD student (Molecular Biology) in Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu, Estonia under the supervision of Dr. Toomas Kivisild and Prof. Richard Villems.
2001–2003	Master of Science (M. Sc.) in Biotechnology from A.P.S University, Rewa, Madhya Pradesh, India.
1998–2001	Bachelor of Science (B.Sc), Honours in Micro-biology from University of Delhi, India.

Work experience:

(Jan 2008 – till present)	As Extraordinary researcher in Estonian Biocentre, Tartu, Estonia (involved in projects related to South Asia).
(2004 – Feb 2007)	As Senior Research Fellow (SRF) in project entitled "Application of molecular fingerprinting in genetic characterization of races and species of lac insects." at Indian Institute of Natural Resins and Gums (IINRG), Namkum, Ranchi, India under the guidance of Dr. Ranganathan Ramani.
2003	M.Sc Dissertation from Centre for Cellular and Molecular Biology (CCMB) Hyderabad, India. Title: Y-chromosome and mtDNA phylogeny among Oraon, Munda, Lohra and Chero tribes of Jharkhand under the supervision of Dr. Kumarasamy Thangaraj and Dr. Lalji Singh (March 2003–August 2003).

Distinctions and Awards:

(i) University topper of M.Sc (Biotechnology) of A.P.S. University, Rewa with 78.3%.
(ii) Merit certificate for International Poster Contest-2000 organized by NCERT in collaboration with UNFPA (United Nations Population Fund).

(iii) First position in Poster presentation at the TUMRI Annual conference 2013.
(iv) 2nd position in Poster contest during XXVIII All India Cultural, Literary and Sports festival- PULSE 2000 in AIIMS (All India Institute of Medical Science).
(v) Qualified GATE 2003 (Graduate Aptitude Test in Engineering) with a percentile score of 85.84.
(vi) Qualified TOEFL (Test for English as a Foreign Language) with a score of 103/120.
(vii) Received ESF Dora T8 fellowship for attending SMBE (Society for Molecular Biology and Evolution) conference in Lyon (4^{th} to 8^{th} July, 2010).
(viii) Won the travel award to attend the 17^{th} ESPCR (European Society of Pigment Cell Research) held in Geneva (11^{th} – 13^{th} Sept, 2012).
(ix) Won the travel award for participation in the EMBO meeting entitled "Human evolution in the genomic era: Origins, populations and phenotypes" held in UK (1^{st} to 4^{th} April, 2014).

Papers and Poster Presentations:

(i) Presented a poster at the TUMRI annual conference entitled "An insight on *SLC24A5* and it's role in human skin color variation" held at the institute (Dec 2013).
(ii) Presented a short talk and poster entitled "An insight on *SLC24A5* gene and it's role in human skin color variation" at the 17^{th} ESPCR (European Society of Pigment Cell Research) held in Geneva (Sept 2012).
(iii) Presented a poster entitled "Investigating the role of *SLC24A5* gene in human skin color variation" in SMBE (Society of Molecular Biology and Evolution) Lyon, France, (July 2010).
(iv) Presented a talk entitled "Study of rs1426654 SNP in *SLC24A5* and it's role in human skin color variation" in EstSHG (Estonian Society of Human Genetics) Vijlandi, Estonia (Oct 2009).
(v) Presented a poster entitled "Investigating the role of *SLC24A5* in human skin color variation in South Asia" in UKIERI (UK-India Education and Research Initiative) conference at LCHES, Cambridge Unuversity, UK (Nov 2008).

Publications:

Basu Mallick C, Iliescu FM, Möls M, Hill S, Tamang R, Chaubey G, Goto R, Ho SYW, Romero IG, Crivellaro F, Hudjashov G, Rai N, Metspalu M, Mascie-Taylor CGN, Pitchappan R, Singh L, Marta Mirazon-Lahr M, Thangaraj K, Villems R, Kivisild T. (2013) The Light Skin Allele of *SLC24A5* in South Asians and Europeans Shares Identity by Descent. PLoS Genet 9(11): e1003912. doi:10.1371/journal.pgen.1003912

Gallego Romero I, **Basu Mallick C**, Liebert A, Crivellaro F, Chaubey G, Thomas M, Villems R, Singh L, Swallow D, Thangaraj K, Mirazon-Lahr M,

Kivisild T. (2012) Herders of Indian and European cattle share their predominant allele for milk tolerance. Mol Biol and Evolution. 2012; 29:249–60.

Metspalu M, Romero IG, Yunusbayev B, Chaubey G, **Mallick CB**, Hudjashov G, Nelis M, Mägi R, Metspalu E, Remm M, Pitchappan R, Singh L, Thangaraj K, Villems R, Kivisild T. (2011) Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. Am J Hum Genet 2011; 89(6):731–744.

Ranjan SK, **Mallick CB**, Saha D., Vidyarth AS, and Ramani R. (2011) Genetic variation among species, races, forms and inbred lines of lac insects belonging to the genus Kerria (Homoptera, Tachardiidae). Genetics and Molecular Biology, 34((3), Jul-Sep), 511–519.

Saha D, Ranjan SK, **Mallick CB**, Vidyarthi AS, Ramani R. (2011) Genetic diversity in lac resin-secreting insects belonging to Kerria spp., as revealed through ISSR markers. Biochemical Systematics and Ecology; 2011; 39: 112–120.

Chaubey G, Metspalu M, Choi Y, Mägi R, Romero IG, Soares P, van Oven M , Behar DM, Rootsi S, Hudjashov G, **Mallick CB**, Karmin M, Nelis M, Parik J, Reddy AG, Metspalu E, van Driem G, Xue Y, Tyler-Smith C, Thangaraj K, Singh L, Remm M, Richards MB, Lahr MM, Kayser M, Villems R, Kivisild T. (2010) Population Genetic Structure in Indian Austroasiatic speakers: The Role of Landscape Barriers and Sex-specific Admixture. Mol Biol and Evolution. 2010; 28:1013–24.

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

Thangaraj K, Chaubey G, Kivisild T, Selvi Rani D, Singh VK, Ismail T, Carvalho-Silva D, Metspalu M, Bhaskar LVKS, Reddy AG, Chandra S, Pande V, Prathap Naidu B, Adarsh N, Verma A, Jyothi IA, **Mallick CB**, Shrivastava N, Devasena R, Kumari B, Singh AK, Dwivedi SKD, Singh S, Rao G, Gupta P, Sonvane V, Kumari K, Basha A, Bhargavi KR, Lalremruata A, Gupta AK, Kaur G, Reddy KK, Rao AP, Villems R, Tyler-Smith C, Singh L. (2008) Maternal footprints of Southeast Asians in North India. Hum Hered. 2008; 66:1–9.

Thangaraj K, Sridhar V, Kivisild T, Reddy AG, Chaubey G, Singh VK, Kaur S, Agarawal P, Rai A, Gupta J, **Mallick CB**, Kumar N, Velavan TP, Suganthan R, Udaykumar D, Kumar R, Mishra R, Khan A, Annapurna C, Singh L. (2005) Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/ insertion polymorphism in Indian populations. Hum Genet. 2005; 116: 507–517.

ELULOOKIRJELDUS

Nimi: Chandana Basu Mallick
Sünniaeg: 25. juuli 1980
Aadress: Evolutsioonilise bioloogia õppetool
Molekulaar- ja Rakubioloogia Instituut
Riia 23b, 51010 Tartu, Eesti
Telefon: +372 737 5005
E-post: cbm2577@gmail.com

Akadeemiline elukäik:

Alates aprill 2007	Doktoriõpe molekulaarbioloogia erialal Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituudis evolutsioonilise bioloogia õppetoolis. Juhendajad: dr. Toomas Kivisild ja prof. Richard Villems.
2001–2003	Biotehnoloogia magistrikraad (*M.Sc.*) A.P.S. Ülikool, Rewa, Madhya Pradesh, India.
1998–2001	Mikrobioloogia bakalaureusekraad (*B.Sc.*) kiitusega, Delhi Ülikool, India.

Töökogemus:

Alates 2008	Erakorraline teadur Eesti Biokeskuses Tartus (kaastööd Lõuna-Aasiaga seotud projektides).
2004 – veeb. 2007	Vanemteadur India Looduslike Vaikude ja Kummide Instituudis Namkum, Ranchi, India projekti „Molekulaarse sõrmejäljemeetodi rakendamine lakiputukate rasside ja liikide määramisel" juures. Juhendaja dr. Ranganathan Ramani.
2003	Magistritöö „Y-kromosoomi ja mtDNA fülogenees Oraoni, Munda, Lohra ja Chero hõimudel Jharkhandis" kaitstud Raku- ja Molekulaarbioloogia Keskuse juures Hyderabadis, Indias. Juhendajad dr. Kumarasamy Thangaraj ja dr. Lalji Singh (märts 2003 – august 2003).

Tunnustused ja auhinnad:

(i) Lõpetanud ülikooli magistriõppe kolme parema lõpetaja seas (78,3%), *M.Sc.* biotehnoloogia erialal A.P.S. Ülikoolis, Rewa, Madhya Pradesh, India;
(ii) Diplom väljapaistva töö eest Rahvusvahelisel Stendiettekannete Võistlusel 2000. a., korraldajaks NCERT koostöös UNFPA'ga (Ühinenud Rahvaste Organisatsiooni Rahvastiku Fond);
(iii) Esimene koht TÜMRI aastakonverentsi stendiettekannete seas 2013. a.;
(iv) Teine koht stendiettekannete seast Kogu India kultuuri, kirjanduse ja spordi festivalil AIIMS'is (*All India Institute of Medical Science*) PULSE 2000, 2000. a.;

(v) Läbis GATE 2003 (*Graduate Aptitude Test in Engineering*) testi tulemusega 85,84 %;
(vi) Läbis TOEFL-i (*Test for English as a Foreign Language*) testi tulemusega 103/120;
(vii) Eesti Teadusfondi DoRa T8 stipendium osalemiseks SMBE (*Society for Molecular Biology and Evolution*) konverentsil Lyonis (4.–8. juuli, 2010. a.);
(viii) Reisitoetus osalemiseks 17. ESPCR (*European Society of Pigment Cell Research*) konverentsil Genfis (11.–13. september, 2012. a.);
(ix) Reisistipendium EMBO konverentsile „Inimese evolutsioon genoomi ajastul – päritolu, populatsioonid ja fenotüübid", mis toimus Inglismaal (1.–4. aprillil 2014).

Suulised ja stendiettekanded konverentsidel:
(i) Stendiettekanne "*SLC24A5* ja selle roll inimese nahavärvi varieerumises" TÜMRI aastakonverentsil Tartus (detsember 2013. a.);
(ii) Lühike suuline ettekanne ja stendiettekanne "*SLC24A5* ja selle roll inimese nahavärvi varieerumises" 17.-l ESPCR (*European Society of Pigment Cell Research*) konverentsil Genfis, Šveitsis (september 2012. a.);
(iii) Stendiettekanne "*SLC24A5* rolli uurimine inimese nahavärvi varieerumises" SMBE (*Society of Molecular Biology and Evolution*) konverentsil Lyonis, Prantsusmaal, (juuli 2010);
(iv) Suuline ettekanne "*SLC24A5* geenis asuva SNP rs1426654 uurimine ja selle roll inimese nahavärvuse vareerumises" Eesti Inimesegeneetika Ühingu aastakonverentsil Viljandis (oktoober 2009. a.);
(v) Stendiettekanne "*SLC24A5* rolli uurimine inimese nahavärvi varieerumises Lõuna-Aasias" UKIERI'i (*UK-India Education and Research Initiative*) konverentsil LCHES'is, Cambridge'i Ülikoolis Inglismaal (november 2008. a.).

Avaldatud teadusartiklid:
Basu Mallick C, Iliescu FM, Möls M, Hill S, Tamang R, Chaubey G, Goto R, Ho SYW, Romero IG, Crivellaro F, Hudjashov G, Rai N, Metspalu M, Mascie-Taylor CGN, Pitchappan R, Singh L, Marta Mirazon-Lahr M, Thangaraj K, Villems R, Kivisild T. (2013) The Light Skin Allele of *SLC24A5* in South Asians and Europeans Shares Identity by Descent. PLoS Genet 9(11): e1003912. doi:10.1371/journal.pgen.1003912
Gallego Romero I, **Basu Mallick C**, Liebert A, Crivellaro F, Chaubey G, Thomas M, Villems R, Singh L, Swallow D, Thangaraj K, Mirazon-Lahr M, Kivisild T. (2012) Herders of Indian and European cattle share their predominant allele for milk tolerance. Mol Biol and Evolution. 2012; 29:249–60.
Metspalu M, Romero IG, Yunusbayev B, Chaubey G, **Mallick CB**, Hudjashov G, Nelis M, Mägi R, Metspalu E, Remm M, Pitchappan R, Singh L, Thangaraj K, Villems R, Kivisild T. (2011) Shared and unique components

of human population structure and genome-wide signals of positive selection in South Asia. Am J Hum Genet 2011; 89(6):731–744.

Ranjan SK, **Mallick CB**, Saha D., Vidyarth AS, and Ramani R. (2011) Genetic variation among species, races, forms and inbred lines of lac insects belonging to the genus Kerria (Homoptera, Tachardiidae). Genetics and Molecular Biology, 34((3), Jul-Sep), 511–519.

Saha D, Ranjan SK, **Mallick CB**, Vidyarthi AS, Ramani R. (2011) Genetic diversity in lac resin-secreting insects belonging to Kerria spp., as revealed through ISSR markers. Biochemical Systematics and Ecology; 2011; 39:112–120.

Chaubey G, Metspalu M, Choi Y, Mägi R, Romero IG, Soares P, van Oven M , Behar DM, Rootsi S, Hudjashov G, **Mallick CB**, Karmin M, Nelis M, Parik J, Reddy AG, Metspalu E, van Driem G, Xue Y, Tyler-Smith C, Thangaraj K, Singh L, Remm M, Richards MB, Lahr MM, Kayser M, Villems R, Kivisild T. (2010) Population Genetic Structure in Indian Austroasiatic speakers: The Role of Landscape Barriers and Sex-specific Admixture. Mol Biol and Evolution. 2010; 28:1013–24.

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

Thangaraj K, Chaubey G, Kivisild T, Selvi Rani D, Singh VK, Ismail T, Carvalho-Silva D, Metspalu M, Bhaskar LVKS, Reddy AG, Chandra S, Pande V, Prathap Naidu B, Adarsh N, Verma A, Jyothi IA, **Mallick CB**, Shrivastava N, Devasena R, Kumari B, Singh AK, Dwivedi SKD, Singh S, Rao G, Gupta P, Sonvane V, Kumari K, Basha A, Bhargavi KR, Lalremruata A, Gupta AK, Kaur G, Reddy KK, Rao AP, Villems R, Tyler-Smith C, Singh L. (2008) Maternal footprints of Southeast Asians in North India. Hum Hered. 2008; 66:1–9.

Thangaraj K, Sridhar V, Kivisild T, Reddy AG, Chaubey G, Singh VK, Kaur S, Agarawal P, Rai A, Gupta J, **Mallick CB**, Kumar N, Velavan TP, Suganthan R, Udaykumar D, Kumar R, Mishra R, Khan A, Annapurna C, Singh L. (2005) Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/ insertion polymorphism in Indian populations. Hum Genet. 2005; 116: 507–517.

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 erotonin. Tartu, 1999, 123 p.
52. **Ebe Sild.** Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva.** Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna.** Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro.** Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane.** Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm.** Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg.** Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild.** The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu.** Studies of the TOL plasmid transcription factor XylS. Tartu 2000. 88 p.

61. **Dina Lepik.** Modulation of viral DNA replication by tumor suppressor protein p53. Tartu 2000. 106 p.
62. **Kai Vellak.** Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu 2000. 122 p.
63. **Jonne Kotta.** Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu 2000. 160 p.
64. **Georg Martin.** Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000. 139 p.
65. **Silvia Sepp.** Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira.** On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000. 96 p.
67. **Priit Zingel.** The role of planktonic ciliates in lake ecosystems. Tartu 2001. 111 p.
68. **Tiit Teder.** Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu 2001. 122 p.
69. **Hannes Kollist.** Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu 2001. 80 p.
70. **Reet Marits.** Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu 2001. 112 p.
71. **Vallo Tilgar.** Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002. 126 p.
72. **Rita Hõrak.** Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002. 108 p.
73. **Liina Eek-Piirsoo.** The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002. 74 p.
74. **Krõõt Aasamaa.** Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002. 110 p.
75. **Nele Ingerpuu.** Bryophyte diversity and vascular plants. Tartu, 2002. 112 p.
76. **Neeme Tõnisson.** Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002. 124 p.
77. **Margus Pensa.** Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003. 110 p.
78. **Asko Lõhmus.** Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003. 168 p.
79. **Viljar Jaks.** p53 — a switch in cellular circuit. Tartu, 2003. 160 p.
80. **Jaana Männik.** Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003. 140 p.
81. **Marek Sammul.** Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003. 159 p

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

103. **Eve Vedler.** Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.
104. **Andres Tover.** Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005. 126 p.
105. **Helen Udras.** Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005. 100 p.
106. **Ave Suija.** Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005. 162 p.
107. **Piret Lõhmus.** Forest lichens and their substrata in Estonia. Tartu, 2005. 162 p.
108. **Inga Lips.** Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005. 156 p.
109. **Kaasik, Krista.** Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005. 121 p.
110. **Juhan Javoiš.** The effects of experience on host acceptance in ovipositing moths. Tartu, 2005. 112 p.
111. **Tiina Sedman.** Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005. 103 p.
112. **Ruth Aguraiuja.** Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005. 112 p.
113. **Riho Teras.** Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005. 106 p.
114. **Mait Metspalu.** Through the course of prehistory in india: tracing the mtDNA trail. Tartu, 2005. 138 p.
115. **Elin Lõhmussaar.** The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006. 124 p.
116. **Priit Kupper.** Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006. 126 p.
117. **Heili Ilves.** Stress-induced transposition of Tn*4652* in *Pseudomonas Putida.* Tartu, 2006. 120 p.
118. **Silja Kuusk.** Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006. 126 p.
119. **Kersti Püssa.** Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006. 90 p.
120. **Lea Tummeleht.** Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006. 94 p.
121. **Toomas Esperk.** Larval instar as a key element of insect growth schedules. Tartu, 2006. 186 p.
122. **Harri Valdmann.** Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.

123. **Priit Jõers.** Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli.** Gata3 and Gata2 in inner ear development. Tartu, 2007. 123 p.
125. **Kai Rünk.** Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007. 143 p.
126. **Aveliina Helm.** Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007. 89 p.
127. **Leho Tedersoo.** Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007. 233 p.
128. **Marko Mägi.** The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007. 135 p.
129. **Valeria Lulla.** Replication strategies and applications of Semliki Forest virus. Tartu, 2007. 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007. 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007. 171 p.
132. **Tatjana Krama.** Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007. 112 p.
133. **Signe Saumaa.** The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida.* Tartu, 2007. 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007. 96 p.
135. **Priit Kilgas.** Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007. 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007. 95 p.
137. **Kärt Padari.** Protein transduction mechanisms of transportans. Tartu, 2008. 128 p.
138. **Siiri-Lii Sandre.** Selective forces on larval colouration in a moth. Tartu, 2008. 125 p.
139. **Ülle Jõgar.** Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008. 99 p.
140. **Lauri Laanisto.** Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008. 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008. 105 p.
142. **Birgot Paavel.** Bio-optical properties of turbid lakes. Tartu, 2008. 175 p.

143. **Kaire Torn.** Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg.** Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd.** Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.
146. **Lauri Saag.** Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.
147. **Ulvi Karu.** Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm.** Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks.** Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu.** Acclimation of stomatal structure and function in tree canopy: effect of light and CO_2 concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš.** Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko.** Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast.** Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats.** Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova.** The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida.* Tartu, 2009, 124 p.
157. **Tsipe Aavik.** Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2009, 112 p.
158. **Kaja Kiiver.** Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja.** Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.
160. **Pille Hallast.** Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB/CGB*) gene clusters: diversity and divergence of young duplicated genes. Tartu, 2009, 168 p.
161. **Ain Vellak.** Spatial and temporal aspects of plant species conservation. Tartu, 2009, 86 p.

162. **Triinu Remmel.** Body size evolution in insects with different colouration strategies: the role of predation risk. Tartu, 2009, 168 p.
163. **Jaana Salujõe.** Zooplankton as the indicator of ecological quality and fish predation in lake ecosystems. Tartu, 2009, 129 p.
164. **Ele Vahtmäe.** Mapping benthic habitat with remote sensing in optically complex coastal environments. Tartu, 2009, 109 p.
165. **Liisa Metsamaa.** Model-based assessment to improve the use of remote sensing in recognition and quantitative mapping of cyanobacteria. Tartu, 2009, 114 p.
166. **Pille Säälik.** The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
167. **Lauri Peil.** Ribosome assembly factors in *Escherichia coli.* Tartu, 2009, 147 p.
168. **Lea Hallik.** Generality and specificity in light harvesting, carbon gain capacity and shade tolerance among plant functional groups. Tartu, 2009, 99 p.
169. **Mariliis Tark.** Mutagenic potential of DNA damage repair and tolerance mechanisms under starvation stress. Tartu, 2009, 191 p.
170. **Riinu Rannap.** Impacts of habitat loss and restoration on amphibian populations. Tartu, 2009, 117 p.
171. **Maarja Adojaan.** Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.
172. **Signe Altmäe.** Genomics and transcriptomics of human induced ovarian folliculogenesis. Tartu, 2010, 179 p.
173. **Triin Suvi.** Mycorrhizal fungi of native and introduced trees in the Seychelles Islands. Tartu, 2010, 107 p.
174. **Velda Lauringson.** Role of suspension feeding in a brackish-water coastal sea. Tartu, 2010, 123 p.
175. **Eero Talts.** Photosynthetic cyclic electron transport – measurement and variably proton-coupled mechanism. Tartu, 2010, 121 p.
176. **Mari Nelis.** Genetic structure of the Estonian population and genetic distance from other populations of European descent. Tartu, 2010, 97 p.
177. **Kaarel Krjutškov.** Arrayed Primer Extension-2 as a multiplex PCR-based method for nucleic acid variation analysis: method and applications. Tartu, 2010, 129 p.
178. **Egle Köster.** Morphological and genetical variation within species complexes: *Anthyllis vulneraria* s. l. and *Alchemilla vulgaris* (coll.). Tartu, 2010, 101 p.
179. **Erki Õunap.** Systematic studies on the subfamily Sterrhinae (Lepidoptera: Geometridae). Tartu, 2010, 111 p.
180. **Merike Jõesaar.** Diversity of key catabolic genes at degradation of phenol and *p*-cresol in pseudomonads. Tartu, 2010, 125 p.
181. **Kristjan Herkül.** Effects of physical disturbance and habitat-modifying species on sediment properties and benthic communities in the northern Baltic Sea. Tartu, 2010, 123 p.

182. **Arto Pulk.** Studies on bacterial ribosomes by chemical modification approaches. Tartu, 2010, 161 p.
183. **Maria Põllupüü.** Ecological relations of cladocerans in a brackish-water ecosystem. Tartu, 2010, 126 p.
184. **Toomas Silla.** Study of the segregation mechanism of the Bovine Papillomavirus Type 1. Tartu, 2010, 188 p.
185. **Gyaneshwer Chaubey.** The demographic history of India: A perspective based on genetic evidence. Tartu, 2010, 184 p.
186. **Katrin Kepp.** Genes involved in cardiovascular traits: detection of genetic variation in Estonian and Czech populations. Tartu, 2010, 164 p.
187. **Virve Sõber.** The role of biotic interactions in plant reproductive performance. Tartu, 2010, 92 p.
188. **Kersti Kangro.** The response of phytoplankton community to the changes in nutrient loading. Tartu, 2010, 144 p.
189. **Joachim M. Gerhold.** Replication and Recombination of mitochondrial DNA in Yeast. Tartu, 2010, 120 p.
190. **Helen Tammert.** Ecological role of physiological and phylogenetic diversity in aquatic bacterial communities. Tartu, 2010, 140 p.
191. **Elle Rajandu.** Factors determining plant and lichen species diversity and composition in Estonian *Calamagrostis* and *Hepatica* site type forests. Tartu, 2010, 123 p.
192. **Paula Ann Kivistik.** ColR-ColS signalling system and transposition of Tn*4652* in the adaptation of *Pseudomonas putida.* Tartu, 2010, 118 p.
193. **Siim Sõber.** Blood pressure genetics: from candidate genes to genome-wide association studies. Tartu, 2011, 120 p.
194. **Kalle Kipper.** Studies on the role of helix 69 of 23S rRNA in the factor-dependent stages of translation initiation, elongation, and termination. Tartu, 2011, 178 p.
195. **Triinu Siibak.** Effect of antibiotics on ribosome assembly is indirect. Tartu, 2011, 134 p.
196. **Tambet Tõnissoo.** Identification and molecular analysis of the role of guanine nucleotide exchange factor RIC-8 in mouse development and neural function. Tartu, 2011, 110 p.
197. **Helin Räägel.** Multiple faces of cell-penetrating peptides – their intracellular trafficking, stability and endosomal escape during protein transduction. Tartu, 2011, 161 p.
198. **Andres Jaanus.** Phytoplankton in Estonian coastal waters – variability, trends and response to environmental pressures. Tartu, 2011, 157 p.
199. **Tiit Nikopensius.** Genetic predisposition to nonsyndromic orofacial clefts. Tartu, 2011, 152 p.
200. **Signe Värv.** Studies on the mechanisms of RNA polymerase II-dependent transcription elongation. Tartu, 2011, 108 p.
201. **Kristjan Välk.** Gene expression profiling and genome-wide association studies of non-small cell lung cancer. Tartu, 2011, 98 p.

202. **Arno Põllumäe.** Spatio-temporal patterns of native and invasive zooplankton species under changing climate and eutrophication conditions. Tartu, 2011, 153 p.
203. **Egle Tammeleht.** Brown bear (*Ursus arctos*) population structure, demographic processes and variations in diet in northern Eurasia. Tartu, 2011, 143 p.
205. **Teele Jairus.** Species composition and host preference among ectomycorrhizal fungi in Australian and African ecosystems. Tartu, 2011, 106 p.
206. **Kessy Abarenkov.** PlutoF – cloud database and computing services supporting biological research. Tartu, 2011, 125 p.
207. **Marina Grigorova.** Fine-scale genetic variation of follicle-stimulating hormone beta-subunit coding gene (*FSHB*) and its association with reproductive health. Tartu, 2011, 184 p.
208. **Anu Tiitsaar.** The effects of predation risk and habitat history on butterfly communities. Tartu, 2011, 97 p.
209. **Elin Sild.** Oxidative defences in immunoecological context: validation and application of assays for nitric oxide production and oxidative burst in a wild passerine. Tartu, 2011, 105 p.
210. **Irja Saar**. The taxonomy and phylogeny of the genera *Cystoderma* and *Cystodermella* (Agaricales, Fungi). Tartu, 2012, 167 p.
211. **Pauli Saag.** Natural variation in plumage bacterial assemblages in two wild breeding passerines. Tartu, 2012, 113 p.
212. **Aleksei Lulla.** Alphaviral nonstructural protease and its polyprotein substrate: arrangements for the perfect marriage. Tartu, 2012, 143 p.
213. **Mari Järve.** Different genetic perspectives on human history in Europe and the Caucasus: the stories told by uniparental and autosomal markers. Tartu, 2012, 119 p.
214. **Ott Scheler**. The application of tmRNA as a marker molecule in bacterial diagnostics using microarray and biosensor technology. Tartu, 2012, 93 p.
215. **Anna Balikova**. Studies on the functions of tumor-associated mucin-like leukosialin (CD43) in human cancer cells. Tartu, 2012, 129 p.
216. **Triinu Kõressaar.** Improvement of PCR primer design for detection of prokaryotic species. Tartu, 2012, 83 p.
217. **Tuul Sepp.** Hematological health state indices of greenfinches: sources of individual variation and responses to immune system manipulation. Tartu, 2012, 117 p.
218. **Rya Ero.** Modifier view of the bacterial ribosome. Tartu, 2012, 146 p.
219. **Mohammad Bahram.** Biogeography of ectomycorrhizal fungi across different spatial scales. Tartu, 2012, 165 p.
220. **Annely Lorents.** Overcoming the plasma membrane barrier: uptake of amphipathic cell-penetrating peptides induces influx of calcium ions and downstream responses. Tartu, 2012, 113 p.

221. **Katrin Männik.** Exploring the genomics of cognitive impairment: whole-genome SNP genotyping experience in Estonian patients and general population. Tartu, 2012, 171 p.
222. **Marko Prous.** Taxonomy and phylogeny of the sawfly genus *Empria* (Hymenoptera, Tenthredinidae). Tartu, 2012, 192 p.
223. **Triinu Visnapuu.** Levansucrases encoded in the genome of *Pseudomonas syringae* pv. tomato DC3000: heterologous expression, biochemical characterization, mutational analysis and spectrum of polymerization products. Tartu, 2012, 160 p.
224. **Nele Tamberg.** Studies on Semliki Forest virus replication and pathogenesis. Tartu, 2012, 109 p.
225. **Tõnu Esko.** Novel applications of SNP array data in the analysis of the genetic structure of Europeans and in genetic association studies. Tartu, 2012, 149 p.
226. **Timo Arula.** Ecology of early life-history stages of herring *Clupea harengus membras* in the northeastern Baltic Sea. Tartu, 2012, 143 p.
227. **Inga Hiiesalu.** Belowground plant diversity and coexistence patterns in grassland ecosystems. Tartu, 2012, 130 p.
228. **Kadri Koorem.** The influence of abiotic and biotic factors on small-scale plant community patterns and regeneration in boreonemoral forest. Tartu, 2012, 114 p.
229. **Liis Andresen.** Regulation of virulence in plant-pathogenic pectobacteria. Tartu, 2012, 122 p.
230. **Kaupo Kohv.** The direct and indirect effects of management on boreal forest structure and field layer vegetation. Tartu, 2012, 124 p.
231. **Mart Jüssi.** Living on an edge: landlocked seals in changing climate. Tartu, 2012, 114 p.
232. **Riina Klais.** Phytoplankton trends in the Baltic Sea. Tartu, 2012, 136 p.
233. **Rauno Veeroja.** Effects of winter weather, population density and timing of reproduction on life-history traits and population dynamics of moose (*Alces alces*) in Estonia. Tartu, 2012, 92 p.
234. **Marju Keis.** Brown bear (*Ursus arctos*) phylogeography in northern Eurasia. Tartu, 2013, 142 p.
235. **Sergei Põlme.** Biogeography and ecology of *alnus*- associated ectomycorrhizal fungi – from regional to global scale. Tartu, 2013, 90 p.
236. **Liis 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.