DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
185

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
185

GYANESHWER CHAUBEY

The demographic history of India: A perspective based on genetic evidence

TARTU UNIVERSITY
PRESS

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

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (in Molecular Biology) on August 9, 2010 by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Opponent: Professor Jaume Bertranpetit, Centro de Investigación Biomédica En Red, Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain.

Commencement: Room No 217, Riia 23, Tartu, on September 20th, at 14:00

The publication of this dissertation is granted by the University of Tartu.

ISSN 1024–6479
ISBN 978–9949–19–445–2 (trükis)
ISBN 978–9949–19–446–9 (PDF)

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 415

TABLE OF CONTENTS

LIST OF TABLE AND FIGURES:

LIST OF ORIGINAL PUBLICATIONS

The current dissertation is based on the following publications referred to in the text by their Roman characters:

I. Thangaraj K*, **Chaubey G***, Singh VK, Vanniarajan A, Thanseem I, Reddy AG, Singh L. (2006). In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. BMC Genomics, 7, 151.

II. **Chaubey G**, Metspalu M, Kivisild T, Villems R. (2007). Peopling of South Asia: investigating the caste-tribe continuum in India. Bioessays, 29, 91–100.

III. **Chaubey G**, Metspalu M, Karmin M, Thangaraj K, Rootsi S, Parik, J, Solnik A, Selvi-Rani D, Singh VK, Reddy AG, Metspalu E, Singh L, Kivisild T, Villems R. (2007). Language shift by indigenous population: a model Genetic study in South Asia. Int J Hum Genet, 8: 41–50 (2008).

IV. **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, 8, 227.

* Equal Contribution

My contributions to the above mentioned articles referred in the present thesis are as follows:

– Had a key role in designing the blueprint of the study, performed most of the experiments, analyzed the data and wrote the paper with the contribution of other co-authors.

ABBREVIATIONS

AMH	anatomically modern human(s)
bp/kbp	base pair/thousand (kilo) base pairs
hg(s)	haplogroup(s)
HVS-I/HVS-II	first/second hypervariable segment of mtDNA
KYA	thousand (kilo-) YBP
LGM	Last Glacial Maximum
MRCA	most recent common ancestor
mtDNA	mitochondrial DNA
np(s)	nucleotide position(s)
PCR	polymerase chain reaction
RFLP	restriction fragment length poymorphism
MYA	million (mega-) YBP
MSY	male specific region of Y-chromosome
NRY	non recombining region of Y-chromosome
OOA	out-of-Africa
SW	southwest
YBP	years before present

Definitions of basic terms, used in current dissertation

Haplotype (= lineage)	mtDNA sequence type with characteristic polymorphisms that encompasses all identical sequences
Haplogroup	*in mtDNA and Y-chromosome phylogenetic studies*: monophyletic cluster of haplotypes (clade) sharing characteristic defining sequence polymorphisms
Founder haplotype	common ancestral haplotype to which all haplotypes under concern coalesce to
Coalescence time	coalescent time estimate to MRCA
Phylogeography	the genealogical study of the spatial distribution of lineages
Star-like phylogeny	phylogeny of a set of sequences that mostly (or all) share their MRCA in the same haplotype; a tree with (virtually) no internal branches
Pleistocene	1.8(6) MYA – 10 KYA
early	1.8 MYA – 730 KYA
middle	730-130 KYA
late	130-10 KYA From the last (Eemian) interglacial till Holocene

1. INTRODUCTION

Last year witnessed a momentous occassion in history, the 200th birth anniversary of Charles Darwin. Despite India's vast biological diversity, Darwin never visited Indian subcontinent but his theory of evolution and natural selection applies worldwide. The publication of Darwin's *Origin of species* in 1859 can be said to be symbolic of the beginning era of evolutionary studies. In his book, Darwin laid out the remarkable evidence demonstrating evolution at different levels in organisms and propounded natural selection as a mechanism that drives diversification and complexity in nature. The *Decent of Man* is famous for his contribution to the hypothesis of sexual selection, but the main goal of the book was to provide evidence that evolutionary principles also apply to humans suggesting that modern humans can be seen as descended from ape like common ancestor. Besides this, it is a common question still prevailing – Who were our ancestors? Where did they orginate from? Which migratory route they followed and how they lived in varying climatic conditions since thousands of years? In the last decade, advances in human population genetics and comparative genomics have played an important role in understanding human genetic diversity and adaptation. Genetic haploid markers such asmitochondrial DNA (mtDNA) and Y-chromosome (NRY-non recombining region of Y-chromosome) have proven to be one of the best tools to learn about the prehistory as well as to test various models in the course of evolution of modern humans. Eventually, the rapid progress in human gene mapping has helped us to understand the complex genetic diseases and population divergence at continental and sub-continental level. It also provided us elementary data for the reconstruction of the phylogenetic history of *Homo sapiens* as well as discovery signatures related to various diseases.

The current dissertation manifests upon the genetic and non-genetic evidence associated with peopling of South Asia. The review of literature provides an overview on the current art of knowledge about peopling and landscape of South Asia. It begins with an elementary knowledge about common haploid DNA markers widely used in phylogeographic studies *i.e.* mtDNA and Y-chromosome. In the subsequent chapters linguistics, archaeology, physical anthropology, cultural anthropology, history, palaeoanthropology and genetic studies along with social structuring is reviewed to visualize the complex prehistory of South Asians. The controversial origin of Austroasiatic speakers is discussed in next chapter. Last part features a glimpse of newcomer *i.e.* whole genome analysis (WGA) approach applied recently to study the fine-scale genetic structure. Finally, the results of present study are discussed with key conclusions.

2. LITERATURE SYNOPSIS

2.1. The fundamental aspects of mtDNA and Y-chromosome, in the light of phylogenetic studies

2.1.1. Structure and importance of mtDNA

Human genome comprises of two independent components: First is a sexually inherited complex nuclear genome, which accounts for 99.9995% of total genetic information while; the second is a simple clonally inherited mitochondrial genome, which covers the remaining minor part. Every human cell has the "second" genome, found in the cell's energy-generating organelle, the mitochondrion. Mitochondria are mostly known for their role as cellular power plants *i.e.*, the production of ATP through oxidative phosphorylation mechanism. Mitochondria generate most of the ATP used to drive cellular reactions that require an input of free energy. Mitochondrial DNA (mtDNA) was first recorded by electron microscopy in 1963 (Nass and Nass, 1963). The human mitochondrial genome is defined by a single type of circular double stranded molecule of 16569 (later 16568 see Andrews et al., 1999) bases. (Anderson et al., 1981; Andrews et al., 1999). Human mtDNA consists of 37 genes, most of which are involved in production of energy and it's storage in the form of ATP. It encodes 13 mRNAs, 22 t-RNAs and 2 rRNAs. *ND1–ND6* and *ND4L* encode seven subunits of complex I (NADH–ubiquinone oxidoreductase). *Cyt b* encodes complex III subunit (ubiquinolcytochrome *c* oxidase reductase). COI–COIII encode for three of the complex IV (cytochrome *c* oxidase, or COX) subunits, and *ATP 6* and *ATP 8* genes encode for two subunits of complex V (ATP synthase). Two ribosomal RNA (rRNA) genes (12S and 16S rRNA) and 22 transfer RNA (tRNA) genes are interspaced between the protein-encoding genes. These provide necessary RNA components for intramitochondrial protein synthesis (for more details about mtDNA structure and function, see Chinnery, 2006). MtDNA has two strands, a guanine rich heavy (H) strand and a cytosine rich light (L) strand. The heavy strand contains 12 of the 13 polypeptide encoding genes, 14 of the 22 tRNA encoding genes and both rRNA encoding genes (Anderson et al., 1981; Andrews et al., 1999). MtDNA has high substitution rate, almost 10 times higher than nuclear DNA (Brown et al., 1979; Haag-Liautard et al., 2008) and relatively higher in non-coding control region.

MtDNA is a pivotal tool in evolutionary, population genetics and molecular ecology. Unlike nuclear genome, mitochondrial genome has no proof reading mechanism. Hence, when mutations occur they are passed on as such to the following generations without any repair or recombination. Besides this, there are many mitochondria in each cell and those with disadvantage are eliminated through intracellular competition. Within a cell, all copies of mtDNA are genetically identical and this genetic state is known as homoplasmy. Mutations and mating generate mtDNA heterogeneity, which is called heteroplasmy. Homo-

plasmy is the basic genetic state of mitochondria, where all the hundreds to thousands of mtDNA copies within a cell or an individual have the same nucleotide-sequence (Birky et al., 1978). It was recently found that vegetative segregation can be used to generate homoplasmic cells by an active process under genetic control (Battersby et al., 2003). Vegetative segregation efficiently removes heteroplasmy at mitosis, leading to the early assumption that individuals generally have only one mtDNA haplotype (homoplasmy) (Birky et al., 1978). It is also facilitated by other mechanisms, such as selection (Rispe and Moran, 2000) and mitochondrial bottleneck (Bergstrom and Pritchard, 1998; Roze et al., 2005). Although, mitochondria have their own genome, it produces only a small proportion of its own proteins, rest of them are encoded by nuclear DNA and imported from cytosol. While each cell contains only two copies of any unique nuclear DNA (one of the paired chromosomes), there are thousands of copies of a given mitochondrial DNA per cell. During fertilization, a sperm cell contributes its nuclear genome, but not its mitochondrial genome to the egg cell (for an alternative argument see White et al. 2008). As a result, the mitochondrial genome of the zygote is determined exclusively by that originally found in the oocyte. The mitochondrial genome is therefore maternally inherited (Giles et al., 1980; Stoneking and Soodyall, 1996) and does not undergo any genetic reshuffling (Olivo et al., 1983; Merriwether et al., 1991; for more details about recombination in mtDNA genome, see Macaulay et al., 1999; Metspalu, 2005; Breton et al., 2007; White et al., 2008; Neiman and Taylor, 2009). Such specific mode of inheritance makes it a unique tool for studying human origin and migration (Cann and Wilson, 1983; Cann et al., 1987; Cann, 1994; Redd et al., 1995; Cann, 2001; Kivisild et al., 1999a,b,2002,2005; Palanichamy et al., 2004; Metspalu et al., 2004; Thangaraj et al., 2005a,b; Macaulay et al., 2005; Friedlaender et al., 2005; Olivieri et al., 2006; Hudjashov et al., 2007; Behar et al., 2008a; Soares et al., 2010; Ref. I–IV).

Everyone carries with them a ‘more or less’ exact copy of mtDNA from their mother and their maternal grandmother and so forth for countless generations. The term ‘more or less exact’ is the key to scientists involved in solving the mystery of human origins. That’s because like all DNA, mtDNA is subject to random mutations over the generations. As these mutations are passed on intact to next generation, they in effect become ‘genetic id’ of family genealogy. For making such genealogy one should have the precise knowledge about the sequence variability of mtDNA. In a mtDNA molecule there are three hypervariable segments one to three- (HVS I-III; nps: 16024–16365, 73–340 and 438–574 respectively). The high number of nucleotide polymorphisms or sequence variants in the three hyper variable fractions of the non-coding control region can allow differentiation among the individuals (Greenberg et al., 1983; Wilson et al., 1993; Lutz et al., 1998). These regions are also collectively known as the Displacement-Loop (D-Loop) region because H-strand replication often pauses a few hundred base pairs after it’s initiation, resulting in a structure consisting of the nascent H-strand associated with its template and displaces third single strand (Chinnery, 2006). The D-loop region is considered to be the

most rapidly evolving art of mtDNA which accumulates base substitutions, insertions or deletions at a rate considerably faster than that of single copy nuclear DNA. In the human D-loop region, the estimates of the rate of substitution were found to range between 2.8 (Cann et al., 1984) to 5 times (Aquadro and Greenberg, 1983) the rate of the rest of the mtDNA.

The phylogeographic structuring of the human mitochondrial DNA variation has provided a genetic approach to study the modern human dispersals throughout the world through the female perspective. Variation between two different mtDNA sequences is greatly due to a mutational event rather than recombinational rearrangements. As time passes, mutations accumulate sequentially and quite often also recurrently along less and less related molecules that combine to form independent lineages known as haplotypes. Most broadly, mtDNA variation has been studied in humans because of its relevance in understanding human evolution, population dispersals, adaptation and in terms of the role that mtDNA mutations play in human diseases (Ballinger et al., 1992; Torroni et al., 1993; Watson et al., 1997; Macaulay et al., 1999; Quintana-Murci et al., 1999; Kivisild et al., 1999a,b,2002,2003a,b; Richards et al., 2000; Endicott et al., 2003; Thangaraj et al., 1999,2003a,2005a; Kong et al., 2004; Bandelt et al., 2007a,b; Palanichamy et al., 2004; Metspalu et al., 2004; Sun et al., 2006,2007; Behar et al., 2008a,b; Chandrasekar et al., 2009; Ref. I–IV).

2.1.2. Structure and importance of Y-Chromosome

The Y-chromosome is male-specific, 60 megabases (Mb) in size linear molecule, but has the least number of genes in comparison to any other chromosome (Jobling and Tyler-Smith, 1995). It is the smallest human chromosome and consists of a short (Yp) and a long (Yq) arm. Of the 27 Y-chromosome genes identified, 9 are located on the Yp and the remaining 18 are on Yq (Skaletsky et al., 2003). The human Y-chromosome plays an important role in sex determination as well as male fertility and it is widely used in infertility, population genetics, forensics and genealogy research. The unique properties of the Y-chromosome are a consequence of the evolution of sex chromosomes in mammals. The sex chromosomes have evolved from a pair of autosomes within the last 300 MYA (Ohno, 1967; Skaletsky et al., 2003; Jobling and Tyler-Smith, 2003; Hughes et al., 2010). A barrier to recombination developed between these 'proto' sex chromosomes, isolating the sex-determining regions and eventually spreading throughout the two homologues. In this process, the original functional elements have been conserved by the X-chromosome, but Y-chromosome has lost almost all traces of the ancestral autosome, including the genes that were once shared with X-chromosome. In the absence of recombination, the accumulation of mutation events led to the degeneration of the Y-chromosome (Skaletsky et al., 2003). Genes known on the human Y-chromosome have been shown to be remnants of genes that were present on the ancient proto-sex chromosome in mammalian genomes and a massive loss of genes from the

proto Y-chromosome was proposed (Charlesworth, 1996; Charlesworth and Charlesworth, 2000). Most of the parts of the Y-chromosome are largely inert and have been shown to have relatively small functional genetic content, concordant with ideas about degeneration of Y-chromosome (Charlesworth, 1996). In other words, genes transposed to the NRY region of the Y-chromosome are vulnerable to degeneration in their later generations. The recombination suppression mechanism in Y-chromosome promotes its gradual degeneration which is a common feature of non-recombining sex chromosomes (Steinemann and Steinemann, 1998; Charlesworth and Charlesworth, 2000). However, more recent comparison of human and chimpanzee Y-chromosomes revealed that the human Y-chromosome has not lost any genes since the divergence of humans and chimpanzees between 6–7 million YBP (Rozen et al., 2003; Hughes et al., 2010).

In addition to sex determination, Y-chromosome plays an important role in spermatogenesis (Sun et al., 2000; Skaletsky et al., 2003; Repping et al., 2002, 2003; Fernandes et al., 2002, 2004, 2006; Thangaraj et al., 2003b). Microdeletion of Y-chromosome, removing the azoospermia factor (*AZF* a, b and c) regions, were found to be responsible for the most frequent genetic cause of spermatogenetic failure, which accounts for 8–15% of the male infertility (Skaletsky et al., 2003; Thangaraj et al., 2003b; Repping et al., 2003; Fernandes et al., 2004). Deletion in the *AZF*c region mapped on the distal Yq11 is the most frequent abnormality associated with spermatogenic failure (Kuroda-Kawaguchi et al., 2001). The *DAZ* (deleted in azoospermia) gene is a strong candidate gene for spermatogenesis, isolated within this region, has a RNA binding domain and is found to be transcribed exclusively in the testicular germ line (Reijo et al., 1995; Saxena et al., 2000). Mutations in *DAZ* gene are associated with 13% of cases of human male infertility and 10%–15% of azoospermic men have shown complete deletion of *DAZ* gene (Poongothai et al., 2009). The *DAZ* gene has an autosomal homolog (*DAZ*L) located on the short arm (p24) of the chromosome 3 (Saxena et al., 1996). Both *DAZ* and *DAZ*L play an important role in germ cell development (Ruggiu et al., 1997; Eberhart et al., 1996; Slee et al., 1999; Reijo et al., 2000).

Numerous polymorphic systems have been used to discuss the issue about human origin. Each has its own shortcomings- all autosomes including X chromosome (as well as pseudoautosomal region of Y-chromosome) have multiple ancestors because of genetic reshuffling. Thus, such unlinked polymorphism systems are often prone to processes which rapidly wipe out unequivocal molecular signature of the past. Mainly three types of polymorphisms (indels, SNPs and microsatellites) are widely used to study the Y-chromosomal phylogeography of world (Hammer et al., 1998; Underhill et al., 2000; Underhill et al., 2001; Jobling and Tyler-Smith, 2003 and references therein) *viz.* indels, SNPs and STRs. Indels are insertions or deletions at particular locations on the chromosome, *e.g.* YAP (Y-chromosome Alu Polymorphism) (Hammer et al., 1998). SNP's are single nucleotide polymorphisms in which a particular nucleotide is changed. They also are known as unique event poly-

morphisms (UEP). STR's are the short sequences of nucleotides (mainly tri or tertanucleotide), which are repeated over and over several times in tandem. Because of the lack of recombination almost all indels, SNPs and micro-satellites on Y chromosome are linked to each other. Such linked poly-morphisms form the basis of haplotype definitions which are most commonly subject to the study of genetic differentiation of human populations. The extant distribution of Y-chromosomal haplotype diversity is being increasingly used as a tool for reconstructing the peopling of world by modern humans, from a male perspective (for reviews, see Underhill et al., 2001; Jobling and Tyler-Smith, 2003; Underhill and Kivisild, 2007). Major advancements in this field derive from (i) the discovery of many single nucleotide polymorphisms (SNPs) and biallelic indels; (ii) the possibility of investigating further level of diversity determined by multi allelic simple tandem repeat loci (STRs).

The biological consequences of male specific chromosome (MSY) evolution in human Y-chromosome accounts for the intense interest of biologists in recent decades (Underhill et al., 2000,2010; Jobling and Tyler-Smith, 2003; Jobling et al., 2004; Repping et al., 2006; Underhill and Kivisild, 2007; Karafet et al., 2008; Xue et al., 2009a; Hughes et al., 2010). UEP (Unique Event Poly-morphism) of Y-chromosome is of greatest importance in evolutionary studies due to combination of several factors *viz.* low mutation rates, non recombining nature, paternal transmission and smaller effective population size which is one-fourth the number of autosomes, thus enhancing genetic drift and founder effect. Evolutionary forces have shaped the behavior and structure of the Y-chromosome in many other ways, influencing features such as repeat content, mutation rate, gene content and haplotype structure. The non recombining region (NRY) region of human Y-chromosome conserve compound haplotype information over time scale spanning prehistory of modern humans (Underhill and Kivisild, 2007 and references therein). Y-chromosome nucleotide changes (SNP) have low mutation rates and therefore are interpreted as unique event polymorphisms (UEP). These markers display geographical localization, tracing back to the origin and thus making them the best tools to measure the extent of male geneflow.

Here, we see that the non-recombining part of Y-chromosome sustains a record of mutational events that have occurred along the paternal lineages throughout the evolution. Similar to mtDNA, the change in Y-chromosome also takes place due to mutation, and doesn't involves complex reshuffling, that occur in other chromosomes, therefore, preserving a simpler record of its history. Y-chromosome carries a wide spectrum of mutations *i.e.* chromosomal changes that occur from generation to generation and which can be used as site or sequence specific markers.

2.1.3. Matrilineal and patrilineal common ancestors

Assuming an ideal population having equal number of females and males, the effective population size varies at different loci. For example, in case of mtDNA and Y-chromosome the effective population size remains quarter of autosomes and there are three copies of chromosome X for every four autosomes. The mechanism of transmission of genetic material to the next generation varies for autosomes and mtDNA/Y chromsome. When DNA is passed from one generation to the next, most of it is reshuffled by random processes (such as recombination-random mixture of chromosomes) which make each person unique from his or her parents as well as from their siblings. In this mode of inheritance, the amount of DNA shared by an ancestor decreases generation by generation according to the Mendelian laws. The size of the fragment shared by more distant relative thus is smaller. Close relatives share larger fragments of DNA from a common ancestor. Some unique DNA, however, remain intact while passing from one generation to other. One of these is mtDNA, transmitted only from mother to child and allocates both men and women to trace their maternal lineages. Second is Y-chromosome (*i.e.* MSY), which is passed only from father to son. Since DNA in Y-chromosome does not recombine with other DNA, it is like a genetic surname that allows men to trace their paternal lineages. Therefore, the power of mtDNA and Y-chromosome analyses is derived from a relatively high mutation rate and the apparent simplicity of inheritance (without recombination), which has simplified modeling of population history in comparison to analysis of nuclear DNA. Unlike mtDNA and MSY region of Y-chromosome, where an individual has a single ancestor, X-chromosome and autosomes both have multiple ancestors and provide a broader picture of an individual's heritage rather than a trail of specific ancestry.

In search of autosomal ancestor(s), using Monto Carlo simulations, it was estimated that all people presently living in this globe share a most recent common ancestor (MRCA) dated to 1415 BC (*i.e.* 3425 years), and any individual who lived 5353 BC (*i.e.* 7363 years) or earlier, can be considered ancestor to everyone on this planet as well as an ancestor to at least one individual living today (Rohde et al., 2004 and references therein). But an individual living now receives little or no real genetic inheritance from the majority of the ancestors because of decreasing smaller proportions of DNA segment from generations to generations in the genealogy (Wiuf and Hein, 1997). Nevertheless, because of its unique pattern of inheritance, mtDNA and Y-chromosome are only informative about a small fraction of ancestors coming directly throughmaternal and paternal line. The numbers of such ancestors are significantly less as compared to autosomal ancestors. But, unlike nuclear DNA, we do inherit the complete intact form of mtDNA and Y-chromosome DNA from those few ancestors.

2.1.4. Rate of mutation in patrilineal and matrilineal inheritance

Genetic variation in any organism is a complex process which arises from a combination of evolutionary forces that consist of successes and failures of genes on a backdrop of neutral variation shaped by genome instability, mutation process and demographic history (Oleksyk et al., 2010). Mutations are elementary units of genetic changes which create the fundamental platform for evolution. For estimating the mutation rate in any biological process, it is important to understand several aspects of medical as well as evolutionary genetics. Estimation of mutation rate offers robust information about the mutational dynamics in any species or population. Besides this, it is essential to determine the process of mutation, whether it is due to random genetic drift or by natural selection (Mishmar et al., 2003; Ruiz-Pesini et al., 2004; Lieberman et al., 2005; Kivisild et al., 2006a; Thangaraj et al., 2008; Charlesworth and Willis, 2009; Amato et al., 2009; Lynch, 2010). Signature of selection depends on type, age and strength of selection events. Natural selection acts in at least three modes: positive, purifying (also called stabilizing or negative, eliminating a damaging allele) and balancing selection (including heterozygote advantage and frequency-dependent selection). Positive selection decreases genetic variation by favoring an advantageous allele, while purifying selection maintains the integrity of functional sequences by eliminating deleterious mutations. Selection in mtDNA has been explored in several studies (Mishmar et al., 2003; Kivisild et al., 2006a; Soares et al., 2009; Endicott et al., 2009; Loogväli et al., 2009). However, there has been little consideration about the possibility of natural selection acting on male specific Y (MSY) chromosomal SNPs. It has been reported that MSY has 16 single-copy X-degenerate genes which are evolutionarily conserved and their pattern of nucleotide variation may help us to get more information about selective importance of MSY and its role in health and disease (Skaletsky et al., 2003; Rozen et al., 2009). The rate of mutation in Y-chromosome is several folds higher than autosomal chromosomes because of its restriction to the male germ line (Jobling et al., 2004). Previously, mutation rate estimation of Y-chromosome was done either by phenotypic observations or by comparisons of homologous sequences among closely related species (Nachman and Crowell, 2000; Kondrashov, 2003). Recently, the use of next-generation sequencing technology has yielded a fair rate of base substitution mutations on human Y-chromosome (for details see Xue et al., 2009a) and potentially added a new insight into human mutation process.

2.1.4.1. Towards the fine tuning of mtDNA clock

The methods of molecular dating of the most recent common ancestor (MRCA) of a set of DNA sequences, either taken from within or between species, is of fundamental importance for the interpretations made from genetic data. Different methods have relied on calibrations of mtDNA mutation rate based on

fossil records, archaeological or climatic evidence, and pedigree data. In the 1980s and 1990s, many studies were mainly carried out by sequencing short fragments of mtDNA, *i.e.* HVS-I and HVS-II. The commonly used HVS-I mutation rate of 1.79×10^{-7} substitutions/site/year (Forster et al., 1996), was deduced from the variation in mtDNA haplogroup A2 lineages in Eskimo and NaDene populations of North America that was assumed to be related to a population expansion following the Younger Dryas event (12,900-11,500 YBP). However, it appeared later that the data set which was used for this calibration included many sequencing artifacts (Saillard et al., 2000) and furthermore, in the light of the complete sequence data, haplogroup A2, as defined by the phylogenetically unstable 16111 position, appears to be polyphyletic (www.phylotree.org). We also know that the mutation rate is not uniform throughout the whole mtDNA genome and the control region shows on an average more than five times higher sequence variation than the coding region (Ingman et al., 2000; for more details see Endicott et al., 2009).

A robust mtDNA substitution rate assessment demands good data quality as well as characterization of the variation of mutation rates among different nucleotide positions. Due to advances in sequencing technology increasingly larger numbers of mtDNA complete genome sequences have become available over the past ten years. These complete sequence data sets have significantly improved the molecular resolution of phylogeographic studies (Palanichamy et al., 2004; Thangaraj et al., 2005a; Sun et al., 2006; Olivieri et al., 2006; Hudjashov et al., 2007; Chandrasekar et al., 2009; www.phylotree.org; Ref. I–IV) as well as provided a source for assessing the variation at different nucleotide positions regarding their functionality (Kivisild et al. 2006a; Soares et al. 2009; Endicott et al. 2009; Loogväli et al., 2009).

The most widely used mutation rates for human mtDNA complete sequences are based on interspecies calibrations assuming certain split times of the humans-chimpanzees clade. Several studies have assumed 6.5 million year old coalescent time of human and chimp mtDNA lineages (Mishmar et al., 2003; Kivisild et al., 2006a). In these calculations, 6 million years has been taken as a consensus estimate for the human–chimp species split (Goodman et al., 1998) whereas additional 500,000 years has been added as for the MRCA of the mtDNA lineages within the ancestral species before the split (Mishmar et al., 2003). More recently, Soares et al. (2009) used a more ancient human-chimp split time in the calibration of mtDNA mutation rate and as a consequence the inferred mutation rate is slower. Besides our lack of knowledge about the precise date of the human and chimpanzee split and the effective population size of the ancestral population, there are other issues on mtDNA clock calibration which have been questionedon and again while molecular dating using mtDNA as a tool. The main concern raised in these studies is about the clock-like behaviour of mtDNA mutations, including a claim for effectively different mutation rates operating at inter and intraspecies scale (Ho et al., 2005), and empirical finding of higher rate of non-synonymous substitutions in Arctic populations as compared to populations from low latitudes. It was manifested

that positive selection for mutations disrupting ATP production would explain their higher frequency in Arctic (Mishmar et al. 2003; Ruiz-Pesini et al. 2004). Further studies have however failed to support the correlation between latitude and the proportion of non-synonymous mutations in mtDNA and found instead the correlation between the age of mtDNA haplogroups and a consistent excess of non-synonymous mutations consistent with time dependency of purifying selection (Kivisild et al. 2006a; Stewart et al., 2008; Soares et al., 2009; Loogväli et al., 2009).

The exponential decay model of Ho et al., (2005) asserted that mutation rates differ at intra and interspecies levels and has since then failed to receive any support (Bandelt et al., 2006; Soares et al., 2009). The recent ML approach calibration (using both coding and control regions to test the time dependency of the mutation rate by Soares et al., (2009) and revised synonymous mutation rate (Loogväli et al., 2009), manifested that saturation at rapidly evolving sites is likely the reason for discrepancies of mutation rate estimates at different time scales.

Thus, we conclude from above chapters, the general structure and possible role of mtDNA and Y-chromosome in inferring the population prehistory, coalescent times and their dispersal events. Due to their uniparental inheritance, they contain less information about an individual's ancestor in comparisonto nuclear DNA. However, since it is passed from generation to generation without recombination, they are the most potential source of information for going back to several generations. The frequency and occurrence of different haplotypes can be used to distinguish populations and to shed light on sub-structures within a population and to study inter and intra population variation (see in later chapters). Some of the chapters, such as recombination in mtDNA and detailed structure of human Y-chromosome and mtDNA have already been discussed extensively in earlier PhD theses from our department (Kivisild, 2000a; Tambets, 2004; Rootsi, 2004; Metspalu, 2005), hence it hasn't been elaborated here.

2.2. South Asia: Some General facts

'South Asia' term allocates to the Southern geographic region of the Asian continent comprising political territories of India, Pakistan, Sri Lanka, Nepal, Bangladesh, Bhutan and the Maldives (Fig. 1a). India, Pakistan and Bangladesh making up the bulk of the subcontinent, Nepal and Bhutan are the Himalayan states while the Indian Ocean Island States are Sri Lanka and the Maldives. The other synonyms of South Asia are Indian subcontinent, *Bharatavarsha*, Southern Asia and SAARC (South Asian association of regional cooperation) countries. It is surrounded by Iranian Plateau from the west, the Himalayas from the north and east and Indian Ocean from south. The Hindu Kush Mountain range that runs through Afghanistan and northern Pakistan is usually considered the northwestern edge of the subcontinent. South Asia ranks among the world's most densely populated regions with 1.6 billion human inhabitants living in the

area of 4.43 million square kilometers (Fig.1b). Physically, South Asia can be divided into three parts-the Himalayas, the Indo-Gangetic-Brahamaputra plains and the Central Indian-Deccan plateau (Fig. 1a). The North is enclosed by the great arc of Himalayas. The mountains that extend west from the Himalayas into the Hindu Kush, enclose the low-lands of Pakistan and Afghanistan. The Himalayan range also dominates the two Himalayan states of Nepal and Bhutan apart from the most parts of northern India. To the South of these ranges lie the alluvial plains of the Indus, the Ganga, and the Brahmaputra embracing the most fertile region of Pakistan, India and Bangladesh. This region offers a living place to more than half of South Asian populations. However, there are some exclusive regions with very thin population density, for example the Thar Desert, the Himalayan Mountains, the Karakoram and the Hindu Kush region. South Asia has land connection in the north, northeast and northwest which facilitates population movements. Thus, the variation in climate and various land forms in South Asia aid in bringing about the most remarkable variation in flora and fauna. Understanding the interactions of society and studying about different populations residing in a region with more than a billion people living on just 2.3% of the global land mass may offer insights on perils and promises of times ahead.

Throughout its prehistory and history, South Asia has been a land of various ethnic groups, languages and cultures. The geography of South Asia is, of course, partly tropical and densely populated. The written phase of history, covering a mere two millennia, includes numerous accounts of invasions into the subcontinent and gives evidence of a multitude of cultural contacts of India with its close or more distant neighbors. There are reasons to believe that such contacts and interactions extend to a far longer prehistoric period of the region. Therefore, it is not surprising to find Indian population genetically and socio-culturally highly diverse as of today. Since the castes, tribes, and religious communities which formulate this subcontinent, are so numerous and diverse that one need to look at the multitude effects of geography and history, effects that have persisted throughout the centuries, to understand the diversity. Present Pakistan and northern India refer to the area of longest Muslim impact, Sri Lanka, Bhutan and northeast India are the Buddhist lands in the subcontinent and Bangladesh differs from West Bengal mainly in its higher number of individuals of Muslim ethnicity. There is a fundamental resemblance in various rituals practiced by people in different regions and therefore, shared ritual patterns can account for some unity among the varieties of the religious beliefs that we can see in South Asia over a long period of time. There is a popular saying which every Indian hears from his/her grandparents that "*Kos Kos par badle paani, teen Kos per baani*" meaning at every one mile taste of water changes and at every three miles dialect and this aptly characterizes the variation dynamics of language and dialect within the subcontinent.

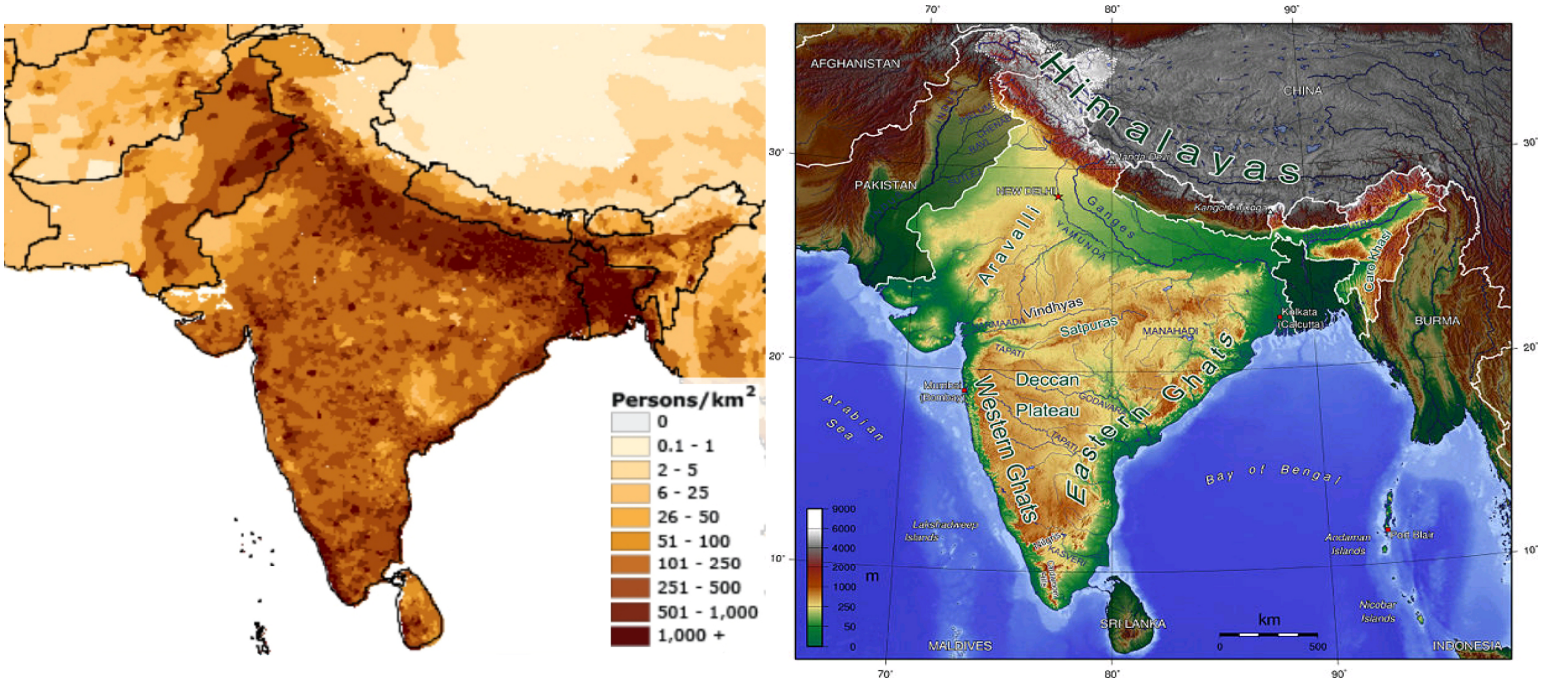

Figure 1. The population density map of South Asia (left pan). The topographical features along with major river water system of South Asia (right pan). Physical division is shown as the upper Himalayas and the associated mountain ranges, the Indus Ganga-Bramhaputra plain extended upto Aravali and Vindhya mountains and the Peninsular Plateau containing Deccan plateau and Eastern-Western Ghats.

The climate of this vast region varies considerably, from tropical monsoon in South to temperate in North. During upliftment of Tibetan plateau and its surrounding ranges, tectonic processes have interacted with climatic changes and with local random events (such as landslides) to determine the development of the major river systems of South Asia. The mountain ranges and Indian Ocean isolate South Asia from rest of Asia. Indian peninsula is covered by the Himalayan range from the North and East and Western Ghats and Eastern Ghats encases it from both the sides providing a strong physical boundary to the peninsula (Fig. 1). Besides this, India has many other mountain ranges, hills and river systems flowing across the country. Some of the hill ranges are very old and are spread across the country. Primarily there are four hill ranges-The Himalayas together with the northeastern hill ranges, the Aravallis, the Central Highlands and the Western Ghats. The Himalayas includes the Shiwaliks, Greater Himalayas, outer Himalayas and the hill ranges like Naga Hills, Garo-Khasi Hills, *etc*. The Central Highlands is composed of many ranges like Vindhya, Satpura, Mahadeo, Rajmahal, *etc*. Further in the South there is Western Ghats. In its lower reaches it bifurcates into Nilgiri Hills, Cardamom Hills, etc. The rivers such as, Ganga, Yamuna, Brahmaputra, Krishna, Kaveri, Indus and Narmada weave a web of water channel all over the subcontinent. The fertile bank of these river systems assisted the subsistence of the populations by providing the elementary needs for a sedentary life.

2.2.1. The major river water systems and their role in population dispersal

Human society has evolved through a complex system of climate and ecological niche. Prevailing records suggest an intimate relationship of adaptations, mitigation and migrations to climate extremes leaving their impact on human society (Núñez et al., 2002). For instance, it has been suggested that increased aridity in Africa led to the eventual rise of arid-adapted hominids and their migration to regions with more conducive climate regimes (deMenocal, 1995). To study the course of human dispersal in South Asia, it is therefore important to learn the role of major river water systems in hominid and modern human settlements. Moreover, they are one of the key determining factors for migration routes and possible two-way dispersal along potential boundaries (Field et al., 2007). The spatial and temporal variations in the rainfall over South Asia has led to denotation of water ‘surplus’ and water scarce river basins across the country (Bandyopadhyay and Perveen, 2002). River courses have also changed significantly and rapidly. Such course changes influenced greatly the sedimentation and subsidence pattern of the basins. The shifting of these rivers produced enormous amount of Holocene aged depositions over the earlier Pleistocene aged ones (Field et al., 2007).

It is evident that presence of water seems to be one the most important reason for all major ancient human civilizations to develop and flourish along

major perennial river systems. For example, civilizations in South Asia, Mesopotamia, and Egypt all developed and flourished along perennially flowing Sarasvati-Indus, Tigris-Euphrates and Nile river. The Indus valley civilization (for detail refers to chapter 2.3.4) is developed in South Asia. It was suggested that during the early Vedic period, the Sarasvati and the Indus were the major river systems (Mishra, 2001). Rakhigarhi in Haryana and Ganweriwala Ther in Cholistan are two of the largest Harappan sites which are more than 100 hectares in size and are comparable to Mohenjodaro, exist on the bank of this holy extinct river Sarasvati (Mishra, 2001). Tumbling down the icy glaciers from the Himalayas (Himadri) in Northwestern Uttarakhand, Sarasvati flowed through Haryana, Southern Punjab, northwestern Rajasthan and eastern Sindh and emptied itself into the Gulf of Kutch (Radhakrishna, 1999; Puri, 2008; Kalyanaraman et al., 2008; Bhadra et al., 2009). This river was mightier than the present day Ganga and Brahmaputra (Valdiya, 2008). In ancient Indian literature this river has been described as a supreme and causative of massive flood in western parts of India (*Rigveda* 2.14.6,6.21.2-9,7.95.1). However, the disappearance of this river is a matter of intensive speculation. The process of disappearance is extensively studied by scholars and the most widely accepted model suggests that it is a case of river piracy by branches of the Ganga and Sindhu rivers (Valdiya, 1996; Puri, 2008; Valdiya, 2008; Kalyanaraman et al., 2008).

Hence, we cerebrate that natural geographical barriers allowed a restricted movement of human populations and their local regional expansion in South Asia. It has been argued that the Thar Desert, the Himalayas, and possibly the Ganges plain would have provided barriers to hominin, population movement according to oxygen isotope stage 4 (OIS) (Field and Lahr, 2005; Field et al., 2007). The geographical barriers not only have affected the transmission of cultural information within the region but also it would have reduced the number of routes for populations dispersing from elsewhere into the region (James and Petraglia, 2005). Such long-term isolation reflects in their genetic makeup and genepool of populations living in several small pockets differentiated with a high degree of separation (Kivisild et al., 2000b,2003a,b; Thangaraj et al., 2005a; Ref. I–IV). R7 and R8 are paradigms of two such regional maternal haplogroups, frequent and restricted to a particular geography (Thangaraj et al., 2009; Ref. IV).

2.3. Peopling of South Asia

2.3.1. Archaeological and palaeoanthropological evidence

Considering the relevance of South Asia in global human evolutionary synthesis, the amount of archaeological studies done till date is less in comparison to genetics. The available distinctive archaeological records in South Asia, though few challenge many of the models and theoretical framework that have emerged through findings made in other continents (Allchin, 2007; Petraglia et al., 2007). The modern human occupation in South Asia has been suggested just after the African exodus, however, there is scarcity of fossil evidence due to unfavorable climate for the process of fossilization (James and Petraglia 2005). The archaeological support for coastal route is elusive, probably due to rise in sea levels after the LGM (last glacial maximum) which would have submerged the settlement (Field et al., 2007). Nevertheless, the presence of lithic components throughout the region indicates occupation of the subcontinent by hominid populations during late Pleistocene period (Kennedy, 2000; Mishra, 2001). The earliest tools, comprising simple cores and flakes, have been reported from the Siwalik hills at Riwat, near Rawalpindi in Pakistan and have been dated to 200 KYA on the basis of magnetic polarity stratigraphy (Dennell et al., 1992). The hominin presence at Narmada river basin appears to have occurred since the Middle Pleistocene (Patnaik et al., 2009). An alternate view however suggests the absence of early humans in the subcontinent during this period (Dennell, 2007). This view still, is not convincing given the evidence of hominin occupation in early Pleistocene time (Dennell and Roebroeks, 2005; James and Petraglia, 2005; Patnaik and Chauhan, 2009). The oldest and only known fossil hominid in India was a partial cranium recovered from Hathnora in the Narmada Basin, Central India (Sonakia, 1984). It's age was calibrated by faunal correlation to ca. 300,000–250,000 YBP (Kennedy, 2000), and taxonomic affinities to *Homo heidelbergensis* (Rightmire, 2001). However, the recent reconstruction dating of calvaria found it considerably younger (160-85 KYA) and put again a big question mark on Narmada man (reviewed in Patnaik et al., 2009; Patnaik and Chauhan, 2009; See also Athreya, 2007). Terminal Pleistocene to mid-Holocene human remains has been discovered throughout South Asia, providing a rich source of information on paleodemography and biocultural adaptations (Kennedy, 2000; Kennedy, 2001). The earliest fossils of modern humans in South Asia have been unearthed in Sri Lanka at Fa Hien Cave dated to ca. 31,000 YBP and Batadomba-lena dated ca. 28,500 YBP (Deraniyagala, 1984). The association of South Asian microlithic industries at sites such as, Fa Hien Cave Sri Lanka, Patne and Jwalapuram in India strengthen the evidence of early presence of modern human in South Asia (Deraniyagala, 1984; James and Petraglia, 2005; Clarkson et al., 2009). The human skeletal remains from this site has been extensively discussed elsewhere (Kennedy, 2000) and so far correspond to the earliest evidence for anatomically modern *Homo sapiens* in South Asia.

The initiation of microlithic technologies provided a highly sophisticated tool for hunting in human prehistoric times. The patterns of microlithic tools, however vary widely among continents *e.g.* Africa (Powell et al., 2009 and references therein), western Europe (Bar-Yosef and Kuhn, 1999), Australia (O'Connell and Allen, 2007), and South Asia (Petraglia et al., 2009), which provides evidence for innovative skills of *Homo sapiens* and also adds another milestone towards independent social convergent evolution. Archeological studies on African microlithic tools suggested emergence of complex technologies offering greater technological flexibility, as well as more sophisticated hunting strategy and planning abilities (Ambrose, 2008; Powell et al., 2009). Similar but independent technological innovation has also been reported in recent archaeological discovery made in South India which signifies that systematic production of small, retouched stone blades was common there by ~28–32 KYA (Petraglia et al., 2009). Such technological innovation can be significantly associated with Late Pleistocene indigenous population expansion in South Asia (Petraglia et al., 2009). The recent discovery from the Jurreru River Valley indicates a rapid and systematic development of microlithic technology across South Asia dating to 35–28 KYA (Petraglia et al., 2009; Clarkson et al., 2009). The discovery of microlithic technology by modern humans in South Asia revolutionized their subsistence, facilitated them to survive in adverse conditions and sustain large population size (who have exploited natural resources in contracting favorable ecological niche just before the onset of LGM) (Petraglia et al., 2009; Clarkson et al., 2009).

Nearly all the rock-shelters in India occupied by the Upper Paleolithic and Mesolithic people, and many others as well, contain rock-paintings depicting a variety of subjects mainly animals or scenes including both human and animals (Mishra, 2001). These rock-paintings are distributed across the length and breadth of the subcontinent. They have been found in Chargul in Northwest Pakistan to Orissa in the East and from Kumaon hill in the North to Kerala in the South. Some of the important rock painting sites are Murhana Pahar in Uttar Pradesh, Bhimbetka, Adamgarh, Lakha Juar in Madhya Pradesh and Kupagallu in Karnataka. The occurrence of hematite pieces found in the occupational debris of Upper Paleolithic and Mesolithic periods conclusively proves that these paintings were made by the occupants of those caves and shelters (Mishra, 2001). Animals were the most favourite subjects of these paintings depicted either alone or in small and large groups shown in various poses. One of the examples is illustration of hunting of the rhinoceros by the Adamgarh group of rock shelters, indicative of large number of people joining hands for the hunt of a bigger animal. The animals are drawn in bold outline and the bodies are sometimes filled in completely or partially with cross-hatching. Examples of all such methods can be seen among the drawings of animals in the caves or rock-shelters at Morhama Pahar in Uttar Pradesh, Bhinbetka (Mishra, 2001), and Adamgarh in Madhya Pradesh. Besides animals, birds and fishes have also been depicted. Depiction of human figures in rock-paintings was also quite common. These are in simple outline forms as well as with hatched body. These paintings

also demostrate common man involved in various activities such as dancing, running, hunting, playing games and battle. A summary of South Asian prehistory based on archaeological findings in chronological order is provided below in table 1. (Mishra, 2001; James and Petraglia, 2005; Allchin, 2007; Patnaik et al., 2009)

Lower Palaeolithic	Sohanian Culture	Named after the river Soan (a tributary of Indus river). Reported at a number of sites in the Siwalik hills in northwest India and Pakistan. Dated interglacial (400,000-300 B.C.) to the end of the Pleistocene period.
	Acheulian culture	First effective colonization of the subcontinent. The remains extensiverly dispersed from the Siwalik hills in the north to areas near Chennai in the south. Acheulian hunter-gatherer populations adapted themselves to a wide variety of ecozones. Sites are particularly densely concentrated and are richer in central India and the southern part of the Eastern Ghats. Climate was essentially semi-arid but it fluctuated several times between cool and dry and warm and wet. In peninsular India, Acheulian artefacts are usually found buried in boulder and pebble gravels of the Chambal, Son, Mahanadi, Narmada, Godavari and Krishna rivers and their tributaries. Tool assemblages comprise choppers, chopping tools, polyhedrons, spheroids, discoids, handaxes, cleavers, scrapers, denticulates, notches, flakes, blades and cores. Dated to lower Pleistocene i.e. between 2.0 and 0.7 million years.
It has been argued that the Acheulian and Sohanian populations inhabited distinct environments, the former occupying the flat surfaces of the Siwalik frontal range and the latter the valleys of the Himalayan flank.		
Middle Palaeolithic		Developed during the upper Pleistocene, A period of intense cold and glaciation in the northern latitudes. In western Europe, the Near East, north Africa and central Asia, the middle Palaeolithic culture is associated with the physical remains of Neanderthal man. Populations occupied the same regions and habitats as the preceding Acheulian populations. Tools were primarily made on flakes and blades. They comprise side scrapers of various types, end scrapers, denticulates, notches, points and borers. Several thermoluminiscence and Th230/ U234 dates from 16R dune profile at Didwana range from 150,000 to 100,000 B.P. Over twenty radiocarbon dates obtained mostly on shell and bone from sites in the northern Deccan and central India range from 40,000 to 10,000 B.P. This shows that middle Palaeolithic assemblages persisted over a long period of time from the terminal middle Pleistocene to the greater part of the upper Pleistocene.
Upper Palaeolithic		Developed during the later part of the upper Pleistocene. Climate was characterized by extreme cold and aridity in the high altitudes and northern latitudes. Human populations were faced with restricted food resources. Vegetation cover over most of the subcontinent thinned out. The discovery of ostrich egg shells at over 40 sites in Rajasthan, Madhya-Pradesh and Maharashtra, dated by 14C, shows that ostrich, a bird adapted to arid climate, was widely distributed in western India during the later part of the upper Pleistocene. Excellent archaeological evidence of this period comes from the Belan and Son valleys. Tool assemblages are essentially characterized by blade and burin tools and show a marked regional diversity with respect to the refinement of techniques and standardization of finished tool forms.Time considered between 50,000 to 10,000 years ago. Recently, it was suggested that the late Pleistocene demographic event coincides with a period of ecological and technological revolution in South Asia. And the emergence of new microlithic technology beginning at 35–30 Kya.
Mesolithic		The subsistence economy continued to be based on hunting and gathering. Remarkable growth in human population, attested by the significantly increased number of sites. First human colonization of the Ganga plains. Technology is primarily based on microliths. The use of bow and arrow for hunting. The first evidence of intentional disposal of the dead. The advent of agriculture.
Neolithic		Agricultural expansion and Animal domestication (also known as Neolithic Revolution). Oldest Neolithic site in South Asia is Mehrgarh. The first animals to be domesticated were dog, cattle, sheep and goat and the first plants to be cultivated were wheat and barley. Rice and Pig's domestication in East Asia. Emergence of sedentry life and and establishment of villages. Copper smelting. Social division. Dated 7000BC to 1400BC in South Asia.
Bronge Age		Dated 3000-1200BC. Addition of tin to copper produced bronze which was a stronger metal than copper. Use of bronze for tools led to the invention of wheel which revolutionized transport and pottery production. Indus-Valley civilization.

Table 1. Summary of prehostorical events, based on Archaeological discoveries. For more detail see the pioneer work of Mishra (2001); James and Petraglia (2005); Petraglia et al. (2009,2010); Patnaik et al. (2009) and Clarkson et al. (2009).

2.3.2. Impact of environment on population dispersal

The monsoon is one of the major weather determinants on earth, and variation in its intensity has a widespread socio-economic impact on the flora and fauna of the subcontinent. For instance, the collapse of civilizations like the Akkadian, classic Maya, Mochica and Tiwanaku were considered to be related with persistent multi-century shifts in climate (deMenocal, 2001). It is, hence, important for South Asia to use paleoclimatic data to examine the variability of the monsoon at long-term scales. The climatic condition of South Asia varies greatly from region to region which can be relevant for the understanding human settlement patterns, barriers in population dispersals, disease and human adaptations. The upper Himalayan region is filled with glaciers. Then, there are dry sand deserts in western part of India and Pakistan, which are characterized by extremes of temperatures ranging from temperatures below 0 °C or even lower in winters to as high as 52 °C in summers. Certain seasons of the year do not have any significant rainfall giving rise to dry and arid conditions. The area between these two vast regions consists of fertile agricultural plains and rivers stretched for hundreds to thousand miles. Having such a varied ecology, South Asia is one of the poorly understood region of the tropics, where rainfall is essentially controlled by the monsoon variability (Gupta et al., 2006). The early palaeoclimatic study in South Asia has been defined by regional pollen zones across the arid and the semi-arid regions and five phases of environmental conditions in northwestern India were proposed (Singh, 1971; Singh et al., 1973). Further studies have been done on environmental fluctuation since Miocene and monsoonal shifts during the Pleistocene and marked seasonal changes in wet and dry periods are thought to have structured hominine, settlement behaviors (Fig. 2) (James and Petraglia, 2005). A number of studies also advocate that the South Asia experienced climatic fluctuations which had significant impact on human population in the region (Bryson and Swain, 1981; Overpeck et al., 1996; Fleitmann et al., 2003; Gupta et al., 2003; Hong and et al., 2003; Staubwasser et al., 2003; Sharma et al., 2004; Gupta et al., 2006; Iyengar, 2009; Patnaik and Chauhan, 2009). Such changing environmental conditions would have affected the settlement and migration patterns of modern humans in the subcontinent. The genetic studies also fortifies demographic expansions at different time scales (Kivisild et al., 1999a,b,2000b,2003a,b; Endicott et al., 2003a; Metspalu et al., 2004; Sengupta et al., 2006; Petraglia et al., 2009; Underhill et al., 2010; Ref. I,IV).

In South Asia, the folklores, rites, festivals and sowing and harvesting of crops are closely associated with the changing seasons. Even in the twenty-first century, the economic planning and the political scenario are largely determined by a good or bad monsoon. Studies dealing with variability of the Indian monsoon has been an active area of research since long due to its heavy impact over agricultural output on which a major proportion of the South Asian population survives (Allchin and Allchin, 1997; Gupta et al., 2003; Sharma et al., 2004; Iyengar, 2009). Changes in the monsoon affect the hominid dispersal and are

preserved in various proxies across the region both on land and in the marine sediments (Caratini et al., 1994). Recent palaeoclimatic, archaeological and historical evidence across regions suggest considerable human dispersals, adaptations, cyclic spatial and demographic reorganization such as abandonment and expansion, and human migrations (Allchin and Allchin, 1997; Staubwasser et al., 2003; Hong and et al., 2003; Gupta et al., 2003; Gupta et al., 2006). It has been argued that the rise and fall of various civilizations in South Asia may have been triggered by climate fluctuations dominated by seasonal changes in the monsoons (Singh, 1971; Allchin and Allchin, 1997; Gaur and Vora, 1999; Staubwasser et al., 2003).

Several meteorological studies show an increase in precipitation in early Holocene and suggested that the early Holocene was an interval of warmer and wetter conditions with intensified SW monsoon (Overpeck et al., 1996; Gupta et al., 2003; Hong et al., 2003). Major South Asian rivers, including the Ganga-Sarasvati-Indus were flowing on their peak during this time (Bryson and Swain, 1981; Radhakrishna, 1999). Two times their present sediment load was being discharged by Ganga-Brahmaputra rivers and widespread peat formation in Tibet during the early Holocene isindicative of increased precipitation than present day SW monsoons over the Indian subcontinent (Goodbred and Kuehl, 2000; Gupta et al., 2003; Hong et al., 2003). The rising of early Holocene northwest monsoon seems to be the major factor which led to the foundation of agriculture as well as early civilization in northwest Indian subcontinent (Singh, 1971; Bryson and Swain, 1981; Goodbred and Kuehl, 2000; Gupta et al., 2003; Hong et al., 2003). The chronology of AMH settlement at Mehrgarh (in present Pakistan) (Jarrige, 1981; Costantini, 1984), is completely in agreement with the peak intensification of the SW monsoon (Overpeck et al., 1996; Gupta et al., 2003; Hong et al., 2003; Fleitmann et al., 2003).

Another earlier prehistoric event which had affected climate as well as potentially the hominine survival and dispersal around the Indian Ocean was the earth's largest volcanic event *i.e.* Mt. Toba supereruption, which happened in Sumatra ~74 thousand YBP (Westgate et al., 1998; Petraglia et al., 2007). Based on pairwise mismatch distributions of mtDNA, it was suggested that modern humans passed through a genetic bottleneck with only a few thousands of survivors whose numbers have expanded then in different continents (Jorde et al., 1998; Ambrose, 2003; Lewin and Foley, 2004). Recent well dated archaeological study in Southern India, based on continuity of middle Paleolithic technology suggested the presence of human before and after the Toba eruption (Petraglia et al., 2007). Though, there is no such fossil record available so far in South Asia, above study assumes that the Out-of-Africa migration had already occurred (Lahr and Foley, 1994), before the Toba eruption and the time coincides with the widely cited genetic coalescence dates (Mishmar et al., 2003). In contrast to this, coalescent analysis of complete mitochondrial DNA sequences, suggests that the Out-of-Africa migration was launched after the Toba eruption (Richards et al., 2006). The recent recalibration of mtDNA molecular clock (Soares et al., 2009; Loogväli et al., 2009), and available genetic information about Out-of-Africa migration also

propose only a single exit of modern human from the horn of Africa (Underhill and Kivisild, 2007). The Toba eruption time doesn't fit now with the recent genetic time frame of Out-of-Africa migration and it therefore needs further revision. Moreover, recently it was shown that the South African archaeological sequence, which was similar to Jwalapuram sequence, is considerably younger than the Jwalapuram, thus questioning the reliability of techno-morphological similarities in identifying long-distance cultural connections and technological dispersal and possible regional innovations (Chauhan, 2010). Therefore, more work is needed across the entire South Asia and in regions closer to the Toba volcano to resolve it.

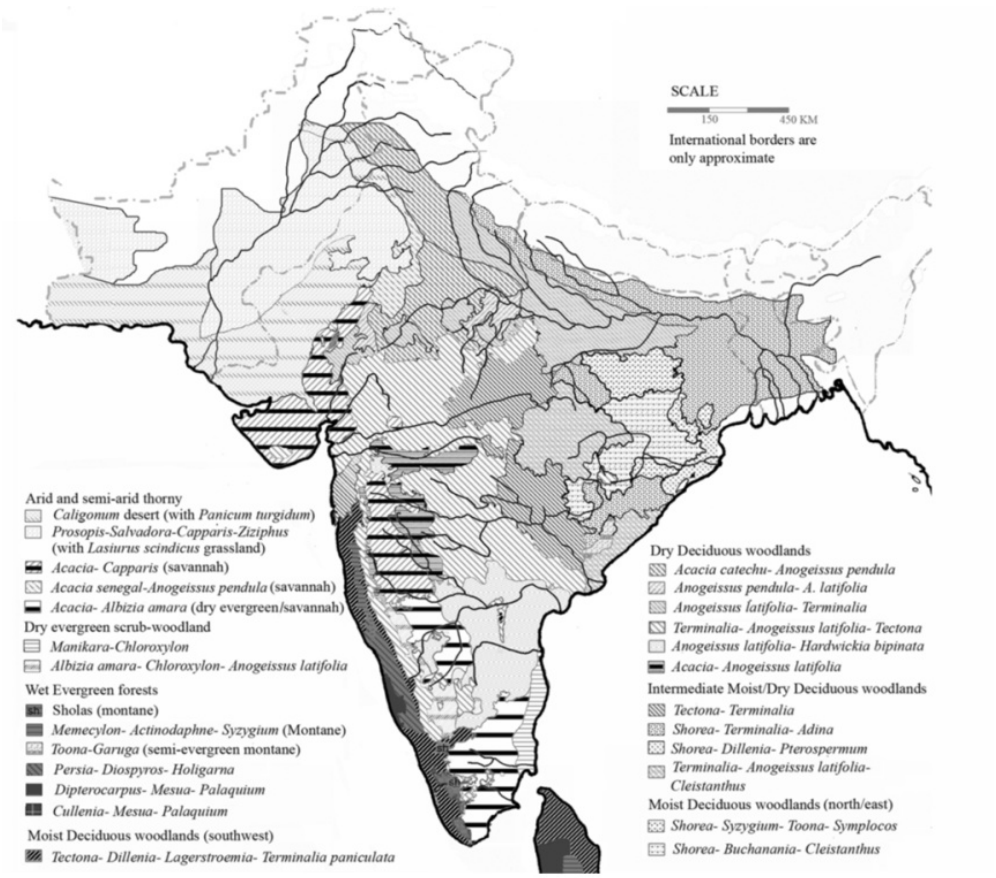

Figure 2. The baseline vegetation map of South Asian palaeoclimate reconstruction for ca. 35–30 KYA. This map is modified from Petraglia et al. (2009), for full legend see Petraglia et al. (2009).

2.3.3. The agricultural expansion and plant/animal domestication

The most fundamental advancement which affected the course of human life was domestication of several plants and animals, resulting in the emergence of more complex societies (Diamond, 2002).The domestication of livestock was a major leap towards the origin of farming and agriculture and henceforth settlement of modern humans. However, the source, time of origin and mode of expansion across South Asia is indeed an intellectual challenge to scholars studying the evolution of culture and genetic history. Adaptation of agriculture practices, allowed humans to shift their diet towards less protein and more starchy food. The significant change in diet must have created various selective pressures over the human genome, acting on the genetic variations of human populations. The agricultural expansion provided a reason to modern humans to survive in a better way and better conditions, led to rapid increase in population and their migration (Diamond, 2002). It has been also suggested that food production conferred enormous advantages to farmers compared to hunter gatherers living outside their homelands and hence triggered outward dispersals of farming populations, bearing their languages and lifestyles intact (Diamond, 1997,2002; Diamond and Bellwood, 2003 and references therein). Furthermore, most hunter gatherer societies were mobile while most food-producing societies were sedentary and could thus accumulate stored food in surplus amount, which was a prerequisite for the development of complex technology, social stratification, centralized states, and professional armies (Mishra, 2001). Genetic findings provide an evidence to Darwinian positive selection on several genetic locus due to the dietary shift in Neolithic period (Coyne and Hoekstra, 2007; Novembre et al., 2007; Perry et al., 2007; Shadan, 2007; Peng et al., 2010). These studies along with other archaeological findings (Richards et al., 2003; Fuller, 2007 and references therein), implied that in early Neolithic period the life standards were quite low, people were less nourished, affected by diseases and the agriculture based diet (mainly starch) was poorer and couldn't provide the essential amino acids which hunter-gatherer diet could, as it was quite diverse and highly rich in proteins (Richards et al., 2003; Eshed et al., 2006; Coyne and Hoekstra, 2007; Perry et al., 2007; Shadan, 2007; Naugler, 2008; Peng et al., 2010). For instance, there are two genetically well-studied paradigms, the copy number variation of amylase gene for starchy food and the regulatory sequence variation of lactase gene for milk consumption (Beja-Pereira et al., 2003; Tishkoff et al., 2007; Perry et al., 2007; Shadan, 2007). Besides this, there were also side effects of agricultural based economy such as the increase the amount of starch and sugars in the diet led to increase in dental cavities (Larsen, 2002). Thus, shifting to agriculture was not very advantageous to the hunter gatherers in terms of health (Fuller, 2007). But in the course of development these societies overcame these pitfalls with the advancement in knowledge and understanding of human physiology and thus helped in developing better food habits. It is significant demographic advantage of

farming in which substantial migration is supposed to have occurred, due to genetic selections and/or language shift (Ref. III).

In the *"Origin of species"* Charles Darwin described his observations on the process of domestication of both flora and fauna by artificial selection, with his theory of *'Natural selection'*. The spectacular advancement in archaeogenetics and archaeobotany has enriched our knowledge on the process of domestication and provided a base to improvise the Darwinian Theory. Domestication is a process where an animal or a plant's life cycle (*e.g.* breeding, habitat *etc.*) is in control of humans. This is the process where phenomenon of artificial selection applies. The domestication of wild organisms (animals and plants) revolves around the principle of not only making them more dependent on humans for their survival but also more productive for themselves. Domestication and the process of food production starting from igniting the fire might have taken a long time for modern humans to disover it and they might have learnt it through a series of observations (Fuller, 2007). Recent archeological studies imply that South Asia has an autochthonous domestication of many local plant species (Fuller, 2006; 2007). Diamond and Bellwood (2003) suggested the incoming of agriculture to South Asia from two different directions - rice cultivation from the East Asia while wheat and barley from the Near East. East Asia has been identified as a homeland of pig domestication, and rice cultivation was imported to South Asia from the Yangtze river basin (China) (Diamond and Bellwood, 2003). The rice cultivation is South Asia has been often correlated with the arrival of Austroasiatic (Munda) speakers (more details in chapter 2.4.4) by several scholars (Diamond and Bellwood, 2003; Diffloth, 2005). However, the contested view of Fuller, (2003) based on archaeobotanical evidence suggests East Indian homeland for rice domestication. Near Eastern region has been identified as a 'cradle of agricultural expansion' and also a source of agricultural and livestock package (combining the cultivation of wheat and barley and the domestication of cattle, sheep and goat) to northwestern part of South Asia (7–9 KYA) (Fuller, 2007) and subsequent expansion to Ganges basin and to the Deccan plateau in Central India (*i.e.* 4 KYA) (Fuller 2003,2006,2007). Seven domestication centers were identified for wheat and barley in the Near Eastern Fertile Crescent region (see Fuller, 2007).

Most of the economically important crops of South Asia have been domesticated during the late Neolithic, *i.e.* 6000–8000 B.C. Recent years have witnessed an expansion of archaeobotanical research in this region, much of it with focus on agricultural origins and plant/animal domestication (Weber, 1998; Fuller, 2003,2004,2006,2007; Saraswat, 1992,1993,2004). Such studies centered mainly on the domestication of food plants, especially staple cereals and to a lesser degree, the pulses. The reconstruction of South Asian prehistory based on archaeobotanical and archaeozoological data warranted a new synthesis on agricultural origins (Thomas and Joglekar, 1994; Saraswat, 2004; Fuller, 2003,2006,2007; Madella and Fuller, 2006). The pioneer work of Fuller (2003,2004,2006,2007), based on archaeobotanical evidence resolved the domestication debate of many crop plants and identified five main domesti-

cation centers in South Asia (Fig. 3). Most of these centres were identified as source of domestication, while many others have obtained agriculture, either through demic or cultural diffusion. Winter as well as summer crops were already a part of the agricultural system at such sites. These crops included indigenous Indian pulses such as horsegram (*Macrotyloma uniflorum*), which is domesticated in Western India and the mungbean (*Vigna radiata*) which could have a western Himalayan origin (Fuller, 2007). The origin of agriculture in Southern India and in Ganges plains probably started independently from West India before the agricultural diffusion from west (Fuller, 2003,2006,2007). Fuller, (2006) suggested Gujarat as a centre for the domestication of local monsoon-adapted crops, such as the little millet (*Panicum sumatrense*) and a species (or two) of *Setaria*. For a list of field crop species and selected fruits with their domestication centers see Fuller (2006,2007).

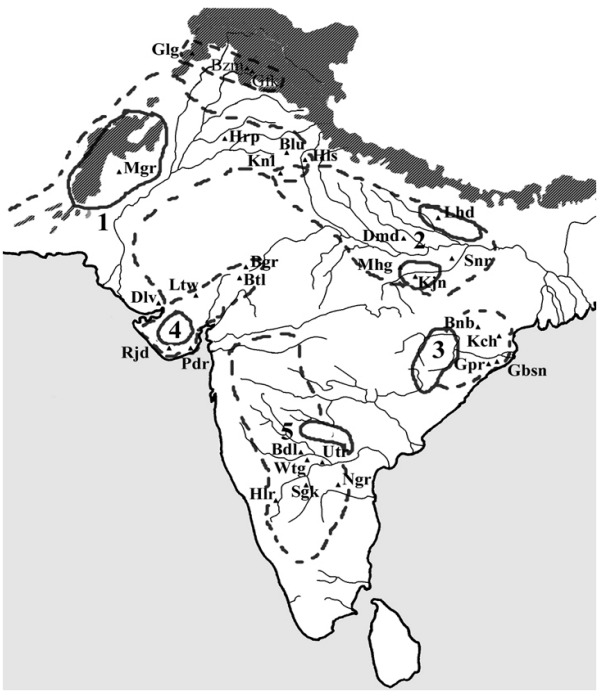

Figure 3. Five main Neolithic domestication centers proposed by Fuller,(2006,2007).

Animal domestication is one of the major topics of archaeological and more recently, genetic investigations (Higham, 1996; Beja-Pereira et al., 2006; Fuller, 2007; Chen et al., 2010). The modern-day domestic cattle has been classified into two groups: zebu (humped) and taurine (humpless), which have their Latin names as *Bos indicus* and *Bos taurus*. Majority of domestic cattle from Europe

and North Eurasia are humpless taurine like (*Bos taurus*), whereas humped zebu cattle predominate in South and Southeast Asia. Zebu is also reported in South China where they are believed to have been introduced from South Asia some 2500 YBP (Higham, 1996). Based on observed morphological differences between cattle depicted in the rock art of South India and in the iconography of Indus-Valley civilizations, South India has been suggested as a secondary center for zebu domestication (Allchin and Allchin, 1997). A clear-cut dichotomy has been described for zebu and taurine mtDNA studies (Loftus et al., 1994; Troy et al., 2001; Beja-Pereira et al., 2003; Beja-Pereira et al., 2006). This dichotomy can only be explained by the presence of two subspecies of wild aurochs before their independent domestication (Troy et al., 2001; Beja-Pereira et al., 2006). There are around thirty well-recognized cattle breeds in India (Vijh, 2000). Many of these breeds are low milk producers; hence used for the production of castrated bulls, which is used in agriculture, carting and transport. The Mehrgarh region is the one of the major well studied Neolithic centre of South Asia, which probably witnessed the first domestication event in the sub-continent (Jarrige, 1981; Costantini, 1984; Allchin and Allchin, 1997; Fuller, 2007). Archaeological studies on Indus-Valley civilization indicate that domestication event of Zebu occurred earlier in the Indus Valley (~8000 YBP) than South India (~5000 YBP) and middle Ganges (~4000 YBP), and parallels geographically to the distribution of probable wild aurochs (Meadow, 2004; Fuller, 2006).

Recently Chen et al. (Chen et al., 2010) analyzed 844 zebu mtDNA control region sequences covering nineteen countries distributed throughout West Asia, South Asia and East Asia including thirty discrete populations. They further grouped them into eight major geographic regional groups. This large-scale study is in agreement with archaeological evidence suggesting the greater Indus Valley (including Rajasthan and present-day Pakistan) as the most likely site for the primary domestication center of Zebu. Based on different phylogeographic distribution of different clades, they further advocated secondary recruitment center of local wild female aurochs into proto-domestic Zebu within Northern India.

Hence, the Neolithic shift towards agriculture was a major change in modern human lifestyle which has been reflected in several archaeological and genomic studies. As a result, the expansion of populations have given rise the diversification of new technology, domestication of new livestocks and shifting of languages either by cultural or demic diffusion (Cordaux et al., 2004a; Sahoo et al., 2006; Ref. III). The detailed study of Zebu domestication provided the evidence of a major step forward in human prehistory, contributing to the emergence of more complex societies. Other domestic animals that remain controversial in terms of their origins in South Asia are water buffalo (Kumar et al., 2007), chicken and pigs, which need a proper genetic scrutiny.

2.3.4. A search for lost civilization: Indus-valley civilization

The discovery of the Harappan Civilization in the early twentieth century was considered to be the most significant archaeological discovery in South Asia, not only because it was one the earliest civilizations of the world, but also because it stretched back the antiquity of well settled life in South Asia by two thousand years at one stroke (Shinde et al., 2009; Gangal et al 2010). This civilization was unique in comparison to the two contemporary civilizations (*i.e.* Mesopotamian and the Egyptian) on the account of its remarkably large coverage and town planning. The matured phase of this civilization covered an area of more than a million square kilometres, attesting it as the largest urbanized civilization of the Bronze Age (Gangal et al. 2010). Indus valley which is situated towards Northwestern part of India has been under the focus of research by archaeologists, geneticists and linguists interested in the continuity and external influences of Indian cultural heritage (Singh, 1971; Kenoyer, 1998; Radhakrishna, 1999; Mishra, 2001; Puri, 2008; Valdiya, 2008; McElreavey and Quintana-Murci, 2005; Fuller, 2006; Fuller, 2007; Gangal et al. 2010). The excavation at Mohanjodaro and Harappa have mapped the key stages of the development of the Indus valley civilization (Singh, 1971; Kenoyer, 1998; Lawler, 2008,2010; Gangal et al. 2010). Since it was at Harappa, where the most characteristic relics of this civilization were first found, it is also known as Harappan civilization. This civilization was blossoming between 2500-1900 B.C. (Lawler, 2008; Gangal et al. 2010). The 1400 settlements discovered so far are distributed in an area of about 1250 thousand square kilometers. It is noteworthy that only ~3% settlements were located on the Indus and it's tributaries while 80% of the settlements were located on the vast plain between the Indus and Ganges which corresponded mainly to the areas around the dried up Sarasvati river, while rest of them were located in Gujarat and Maharashtra (Singh, 1971; Radhakrishna, 1999; Gaur and Vora, 1999; Mishra, 2001; Gangal et al. 2010). The genetic study of contemporary Indus populations revealed a mosaic composition of different components from West Eurasia, South Asia and East Asia (Quintana-Murci et al., 2004; McElreavey and Quintana-Murci, 2005). Population at East of Indus shows prevalent South Asian maternal lineages, while populations located West of Indus exhibit mainly Western Eurasian lineages (Quintana-Murci et al., 2004).

The work carried out at the site of Mehrgah in Baluchistan (also known as greater Indus valley) has already demonstrated the origins of this culture, which was gradual from the modest beginning of settled life around 7500 BC (Jarrige, 1981; Costantini, 1984). The archaeological data suggest the emergence of social behavior in the Indus basin near Mehrgarh (now in Pakistan) and it's surrounding areas in the Kachhi plain on the bank of the Bolan river (Jarrige, 1981; Costantini, 1984). Mehrgarh was considered to be one of the centres for early civilization and for the shift from hunting and gathering societies to those with settled agriculture and domesticated animals (Allchin and Allchin, 1997).

The favorable climate and abundance of vegetation, including wild cereals, presence of wild animals, and availability of water might have attracted the hunter gatherer communities for settlement and provided the basic needs for commencement of agriculture. Thus, Mehrgarh provides an important evidence for the change from hunting-gathering and pastoralism to a subsistence economy centered around settled agriculture and domestication of wild animals (Jarrige, 1981; Costantini, 1984; Allchin and Allchin, 1997). The agricultural economy of the civilization was thought to be primarily based on the cultivation of wheat and barley and of millets like *jowar* (*Sorghum bicolor*), *bajra* (*Pennisetum typhoideum*), *ragi* (*Eleusine coracana*), little millet (*Panicum miliare*) and Italian millet (*Setaria italica*) in the semi-arid region of Gujarat (Mishra, 2001). One of the recent study focusing on the discovery of silk in the Indus Valley pushes back the earliest date of silk outside of China by a millennium and is roughly contemporaneous with the earliest evidence for silk from within China (Good et al., 2008).

The language of Harappans, however, has been much of a speculation. There are several hypotheses about the nature of this unknown language. Parpola considered it as an Indo-Aryan, based on the continuity with North India where Indo-Aryan language is prevalent (Parpola, 1988; see also Renfrew, 1991). Further, it was suggested that the proto-Elamo-Dravidian language could have spread Eastwards with the movement of farmers from this region to the Indus Valley and the Indian subcontinent (McAlpin, 1981; Cavalli-Sforza, 1996; Renfrew, 1996). But this theory weakens after the recent archeaobotanical and genetic research suggesting indigenous origin of Dravidians (Fuller, 2003; Sengupta et al., 2006). Alternatively to Elamo-Dravidian hypothesis, the language of Harappan was suggested as an extinct para-munda, similar to Austro-asiatic (Khasi) language (Witzel, 2007). The Harappan writing remains undeciphered and the language of the Bronze Age Indus Valley is still one of the great mysteries. Several attempts have been made to decipher the Harappan script but none of them have been widely accepted. It was suggested first as a non-linguistic symbol (Farmer et al., 2004). But more recently, Rao et al. (2009) compared the pattern of symbols to various linguistic scripts and nonlinguistic systems based on a computational algorithm and revealed that the pattern of Indus script is closer to Sumerian and Old Tamil script and that of spoken words, supporting the denotation as a language which was so far unknown.

The abrupt fall of this civilization is a highly debated issue besides the language. Over the past several decades, many theories have been explained behind the sudden disappearance of this well flourished civilization. These varied explanations include reduction in rainfall, foreign invasions, tectonic rise of the land leading to blocking of the flow of the Indus river and eventual migration of the people from Mohenjodaro, environmental degradation due to excessive use of soil and plant resources and hydrological changes (Saraswat, 1993; Mishra, 2001; Tripathi et al., 2004; Gupta et al., 2006; Lawler, 2008; MacDonald, 2009; Gangal et al. 2010 and references therein). From the perspectives of paleolimnology and paleoclimatology, perhaps the most intriguing debates revolve around the

hypothesis of the drought theory based upon paleoecological data from lake sediments arose from the work of Singh et al. (1971,1973), at Lake Didwana in the Thar Desert of Western India. Based on palynological, evidence, Singh (1971) suggested that increase in the monsoon rainfall directly contributed to the rise of the Indus civilization and a decrease in precipitation contributed to it's fall, but this claim has been criticized by many archaeologists (Madella and Fuller, 2006) and references therein. The drying up of river Sarasvati and shifting of whole Indus-Valley population towards Ganga plains is another theory given behind the rapid fall of this civilization (Mishra, 2001; Tripathi et al., 2004; Gupta et al., 2006). Madella and Fuller, (2006) have considered strategic local shifts in agriculture in response to prolonged droughts in 2200BC which have contributed to the de-urbanization process and restructuring of human communities. However, none of them have been really conclusive.

Henceforth, we conclude that the Indus civilization needs a deeper multi-disciplinary approach to uncover the truth. Systematic genetic research of contemporary populations near to the Sarasvati basin is required. There have not been many multi-disciplinary approaches to the Harappan archaeology in India and Pakistan as well. Archaeological research on the Harappan culture needs support and active participation of scholars from various other fields too.

2.3.5. Major language groups

In tracing our origins, various disciplines provide different insights. Language, genes and culture are three distinct and, in principle, independent entities which demonstrate different spread patterns in space and time. There are two categories of information in the universe, first is genetic information what we know as nature, and second is the extra-genetic information, what we call culture. The spread of language or any other cultural element can occur together with total replacement of genes (demic diffusion) or without (cultural diffusion) any exchange of genetic material between the human groups that interact with each other. The models of demic and cultural diffusion can be considered as two extremes whereas in reality the spread of any cultural element may be accompanied by some but not total exchange of genes. The reconstruction of the past is further complicated by the fact that in time many languages and cultures have become extinct, *e.g.* the Hittite language which forms one of the most basic branch of the Indo-European language tree. Human genetic diversity observed in South Asia, as it is discussed below, is second only to that of Africa. High region specific genetic diversity implies an early settlement and demographic growth in South Asia after the African exodus of anatomically modern humans (AMH) in Late Pleistocene. In contrary, linguistic diversity in India has been thought to derive itself from more recent population movements and episodes of contact (Diamond and Bellwood, 2003). With the exception of Dravidian, which has likely local origins in India (Fuller, 2003), the presence of other language families (Indo-European, Austroasiatic, Tibeto-Burman) in India is likely due to recent introgression.

The Indo-European language tree has Hittite, Tocharian, Armenian, and Greek as its most basic branches (Gray and Atkinson, 2003). This branching pattern supports the model of Anatolian origins this language family (Gimbutas, 1970; Mallory, 1989; Renfrew and Boyle, 2000; Gray and Atkinson, 2003). The Indo-Iranian branch of this tree has been dated at 4,600 years whereas all Indian languages in this group coalesce to a common ancestral language at a time depth of only 2,900 years assuming a fairly constant rate of language evolution (Gray and Atkinson, 2003). Dravidians are restricted to South India with some exceptions like Kurukh, Gondi and Malto branches in Central, Eastern India-Nepal and Brahui in Pakistan. The Sino-Tibetan speakers are dispersed throughout the Himalayan belt and mainly concentrated in the Northeastern parts of India. The Austroasiatic people are mainly dispersed in "heartland" of India *i.e.* central and eastern parts. In all, 676 language groups have been identified in South Asia (Ethnologue). A detailed view of different language groups in South Asia has been illustrated in fig.4.

Figure 4. Geographical distribution of different language groups in South Asia (modified from Van Driem 2001)

It was not only the lexical comparison but evaluation of whole languages as organic systems of grammatical regularities, which allowed Marcus van Boxhorn in 1643 to recognize the Indo-European language family. Indo-European is world's largest language family in terms of number of speakers. The name indicates the range of its expansion from Indian subcontinent to Europe. For the first time it was noted by Sir William Jones, who was a judge in Presidency of Bengal and one of the founders of Asiatic Society in Calcutta and he had proclaimed that Sanskrit was related to Greek and Latin. Though, it was not him who discovered it first, but this announcement made him immortal in the linguistic history (for a full story see van Driem, 2001). About half a century later this work was catalyzed by another distinguished orientalist Max Muller, who postulated the "Aryan Invasion theory" which has been one of the most controversial South Asian topics for over a century. However, it should be noted that it still remains just as a theory. To date, no strong archaeological or genetic evidence has proven the Aryan invasion theory to be a fact (Endicott et al., 2007 and references therein). The Indo-European family tree is split into major branches *viz.* Anatolian, Greek, Armenian, Albanian, Tocharian, Indo-Iranian, Italo-Celtic, Balto-Slavic and Germanic (van Driem, 2001). Notably, Armenian represents a distinct branch of this family and is spoken in northeast Asia Minor and south of the Caucasus (Renfrew, 1991). It's closest living relative within the family is probably the geographically distant Albanian language (van Driem, 2001). Many hypotheses have been proposed on the origin and dispersal of this language but so far none is widely accepted.

Few studies point homeland of Indo European language in the Pontic-Caspian region i.e above the Black and Caspian Seas, which today forms southern Russia and southern Ukraine (Gimbutas, 1970). Alternate view argues that Anatolia (what is today central and south-eastern Turkey) is the cradle of Indo-Europeans and the point of their dispersal to their historical habitats (Renfrew, 1996; Renfrew, 2000). This view presumes that the Hittites and other people who spoke related languages (Luvian, Pallaic) are autochthonous. Historians, archaeologists and expert Hittitologists disagree and regard the Hittites intrusive to Anatolia (Gurney, 1990; Dunstan, 1998; Puhvel, 2004). The Anatolian languages Hittite, Luvian and Palaic are recorded in cuneiform writing on over 25000 clay tablets dating to the period between 1650 and 1200 BC (van Driem, 2001). Some of the views based one early *Rigvedic* hymns place *Sapta-sindhu* region as a cradle of this language group (Kazanas, 1999; Feuerstein et al., 2001). The common model of Indo-European speakers in South Asia explains their entry from the northwestern part and their rapid expansion into the subcontinent assimilating large number of Munda and Dravidian languages. The spread of this family has also been associated with domestication of horse by 4500 BC as well as by invention of spoked wheels adopted and associated with the ancient Indo-European pastoralists to enhance their mobility,which enabled them to exploit the open steppe and facilitate their expansion throughout Eurasia (Gadgil, 1997; Witzel, 2005). However, archeological evidence points out that there is no evidence of horses antedating, 2000 BC in Indian

archaeological records (Parpola, 1988). Moreover, the recent extensive genetic studies (see in later chapter) dealing with haploid markers (mtDNA and Y-chromosome) doesn't support this theory (Kivisild et al., 1999a,b, 2003a; Metspalu et al., 2004; Sengupta et al., 2006; Endicott et al., 2007; Ref. II).

Dravidian languages were first recognized as an independent family in year 1816 by Francis Ellis. In 1856, the Dravidian scholar Robert Caldwell proposed a relationship between the Dravidian languages and the language of then freshly deciphered from Elamite inscriptions at Behistun and also gave the name *Dravida* to this language family. Dravidian scholar David McAlpin (McAlpin, 1981) analyzed large amount of data and provided the evidence in support of the Elamo-Dravidian theory. His language reconstruction suggested that proto-Dravidian speakers migrated to Indus Valley and Indian subcontinent from Near East with the farming dispersal (Cavalli-Sforza, 1996; Renfrew, 1996). This hypothesis was supported by many linguistic, archaeological and genetic studies (Cavalli-Sforza et al., 1994; Diamond and Bellwood, 2003; Quintana-Murci et al., 2001,2004; McElreavey and Quintana-Murci, 2005). However, the presence of North Dravidian group (*i.e.* Brahui) in Pakistan has raised a question against this theory. This putative relation was criticized on the basis of lexical affinities which were supposed to be borrowed from each other (Blazek and Boisson, 1992; Blazek, 1999). Well argued archaeobotanical evidence against Elamite link was put forward by Fuller (2003) suggesting indigenous origin of Dravidian language and Brahui as a westward expansion of this family. The bidirectional movements of Dravidians were proposed from their homeland (*i.e.* South India) – Eastward migration towards East of India and Nepal (presently represented by Kurukh and Malto branches), and another 'Out-of-India' migration associated with animal herding and wild food processing, through Saurashtra (Gujarat) or Rajasthan (represented by Brahui population in Balochistan). This view has also been supported by recent Y-chromosomal study (Sengupta et al., 2006).

In South Asia, Sino-Tibetan language family is unambiguously considered to be arrived from East and is mainly present in Nepal, Bhutan and Northeastern part of India and also in Pakistan where the Balti population from the Karakoram Mountains also speaks a Tibeto-Burman branch (Fig. 4). This is the second largest language family in the world after Indo-European. The Sino-Tibetan language originated in China and spread from the Yellow river of China into Burma and the greater Himalayan region (van Driem, 2001). The overwhelming majority of languages spoken in the greater Himalayan region belong to this language family. Though, the greater Himalayan region also comprises of at least six other language families, *i.e.* Indo-European, Dravidian, Austroasiatic, Altaic and the distinct language isolates represented by the ethnolinguistic relict groups Kusunda and Burushaski (van Driem, 2001). The internal classification of Sino-Tibetan is still controversial (Blench, 2008) and needs a thorough vocabulary and lexical reconstruction. Genetic studies on limited populations of this group however, suggest their recent migration from

East to Indian subcontinent (Cordaux et al., 2004b; Metspalu et al., 2004; Reddy et al., 2007).

The study of lexicon as well as grammar enabled Francis Mason in 1854 to study the relationship between the Munda of India and the Mon-Khmer languages of Southeast Asia, to evaluate their relationship and also differentiate them from Sino-Tibetans, and propose the Austroasiatic language family. There are more than 150 Austroasiatic groups scattered all over ranging from the heartland of the Indian subcontinent to Vietnam in the East and the Malay Peninsula in the South. In South Asia distribution of Munda languages is patchy (see Fig. 4). The tangled origin of this language group has been discussed in the subsequent chapter. It was proposed that Munda has undergone changes in their word order and also share some linguistic features with Dravidian and Indo-European families, which point to its long-term bilingualism with non Austro-asiatic languages (Blench, 2008).

2.3.6. Enigmatic populations

Linguistic families as well as language isolates bear witness to older populations of distinct ethnolinguistic stock. Hence, it is important to study the wide aspect of these groups. Though, each and every population of South Asia is unique in terms of their ways of living and specific rituals, there are few relic populations living apart from those thousands of tribal and caste populations. Island populations, for example, Andaman Islanders (Onge, Jarwa, Sentinles, Andamanese, Nicobarese and Shompen) were studied by linguists, archaeologists and geneticists (Cooper, 2002; Endicott et al., 2003a,b,2006; Thangaraj et al., 2003a,2005a,b,2006a,b,2007; Abbi, 2006; Palanichamy et al., 2006; Mellars, 2006a,b; Barik et al., 2008; Reich et al., 2009). Mainland language isolates, for example Nehali or Nahali in Central India, Kusunda in Central Nepal, Vedda in Sri Lanka, Burushaski in Pakistan and Gongduk language of Bhutan can prove to be quite interesting populations for study. Similarly, some tribes such as Didayi, Bonda and Mankiridia of Eastern India (Orissa state) and at some extant Abujhmaria of Chhattisgarh, Malani of Himachal Pradesh and Chola of Kerala also exhibit some distinct attributes in their culture when compared to other local or adjoining tribes. It was suggested that the Burúsho of Gilgit, Vedda of Ceylon, Nehali of the Gawilgarh hills and Kusunda of Nepal are direct linguistic descendants of population groups which were possibly the oldest inhabitants of the Indian subcontinent (van Driem, 2001). They are different from their contemporary populations either in their way of living, language or phenotypic character. These groups didn't have any contact and remained isolated from contemporary populations since their existence, therefore, their genetic analysis can be extremely valuable in understanding genetic basis of several Mendelian disorders.

The highly enigmatic relict population in South Asia arethe Andaman Islanders, who used to speak more than a dozen different tongues (Abbi, 2006).

The languages of the Andaman islanders have been classified into two, the Little Andaman and the Great Andaman group. The Little Andaman group trifurcates in to three branches *i.e.* Onge, (spoken on Little Andaman), Sentinelese, (spoken on North Sentinel Island) and Jarawa (spoken in the interior of South Andaman). Great Andaman group bifurcates itself in to Northern and Southern groups (Abbi, 2006). There is a high degree of dissimilarity between the languages of Little and Great Andaman (Radcliffe-Brown, 1948). None of the linguistic evidences support a relationship between the Andamanese languages and any other known linguistic group, so far (the results of genetic studies related with them have been summarized in chapter 2.4.3.2.). Another hunter-gatherer population inhabiting Great Nicobar Islands is known as Shompen. Until recently, they were grouped with Nicobarese in Mon-Khmer branch of Austroasiatic language (Ruhlen, 1991) but Blench (2007) classified them as language isolate based on their relative phonology. Blench (2007) suggested that Shompen might represent a relic of early human expansion around the rim of the Indian Ocean. Although, the recent genetic studies on contemporary Shompen population, though at low resolution (Trivedi et al., 2006) found a similar genepool as Nicobarese (Prasad et al., 2001; Thangaraj et al., 2003a,2005a,b) and suggested a Southeast Asian genetic link. East Asian specific mtDNA haplogroups B5a and R12 (now R22) are two common maternal founders while their paternal lineages are composed of Austroasiatic specific haplogroup O2a (M95) (Trivedi et al., 2006). The mtDNA as well as Y-chromosomal data have been consistent with the affinity of Shompen with Southeast Asian populations (See Fig. 5 from Trivedi et al., 2006 for contrasting results of haploid and diploid markers), but due to low molecular resolution, the time of MRCA of Shompen with Nicobarese is not known. Thus, the exact origin of Shompen is largely unknown and common effort is needed from all the disciplines of modern human studies.

The Himalayan kingdom of Nepal is incredibly rich and complex in cultural as well as linguistic diversity. This diversity is the result of coexistence of several ethnic groups. Kusunda is one of the intriguing ethnic groups whose language, culture and genetic architecture would be interesting to study. Kusunda is the linguistic descendants of an aboriginal population who inhabited the Himalayan region before the entry of the Sino-Tibetan and Indo-Europeans (van Driem, 2001). They are also known as "*Ban Raja*" (People of forest). The total number of Kusunda speakers is very small, in 2001 Census of Nepal there were only 164 Kusundas reported. They call themselves as descendents of King *Kusa* (son of goddess Sita). The reference of *Kusa* comes in Hindu sacred epic *Ramayana* written by Sage *Valmiki*. A linguistic study suggests their link with Ocenian as well as Andaman Islanders (Whitehouse et al., 2004). However, in the light of current available genetic evidence it is very hard to believe on this putative link. The molecular genetic evidence is needed to determine the origin of Kusunda and their relation with contemporary populations.

Vedda is another isolated small hunter gatherer tribe living in Northwest province of Sri Lanka. They are called as aboriginal people of Sri Lanka and are

most distinct from Sinhalese in morphology. This relic population was suggested to represent the indigenous population of the entire Indian subcontinent (van Driem, 2001). The first genetic study of Vedda along with other Asian populations suggested their long period of isolation (Harihara et al., 1988). However, the analysis of alpha-2-HS-glycoprotein allele frequencies supports the view that the Veddas are biologically most closely related to the Sinhalese (Umetsu et al., 1989). A study on 9bp deletion of mtDNA on Asian populations didn't observe any 9bp deleted individual in Vedda (Harihara et al., 1992). Till date, a high resolution genetic data is not available from this population.

The Nahali (also Nehali or Nihali) is a language isolate spoken in West-Central India (in Madhya Pradesh and Maharashtra) (Lewis, 2009). They are found in the Gawilgarh Hills south of the Tapi River in Nimar and Ellichpur Districts of Madhya Pradesh and in the Amravati and Buldana districts of Maharashtra, across the Vindhya and Satpura hilly ranges, which separate the Deccan Plateau from the Gangetic plain of North India. Besides minor notes by historians (van Driem, 2001) and the Anthropological Survey of India (Singh, 1997), many anthropologists mention Nahali as people having traditionally lived in the Satpura and Vindhya forests as robbers of settled neighbors. They were probably foragers who, after the strengthening of regulations by the British rulers (1857), slowly moved as workers in the fields of agriculturalists, settling down in permanent villages at the margins of the forests. Even today, they commonly live in mixed settlements of varied ethnicities (commonly Korku, Bhil and Gond), where they usually cluster in specific quarters, with no land property of their own, subsisting on wage labour. The assimilation of Nahalis into Kokru initiated with the develpoment of a socio-economic symbiotic relationship which further bolstered, since the Nahalis lost the benefits granted to Scheduled tribal populations since 1991 (Census 1991,2001). Korku language is an outlier too, being the only and Westernmost Austroasiatic idiom spoken in Central India, belonging to the North-Munda branch.

The available data on the Nahali language is limited, with very little known about grammar (Witzel, 1999). On the basis of word lists compiled during Grierson's Linguistic Survey of India (1903-1928), linguists (Bengston, 1996; Mundlay, 1996; Whitehouse, 1997) debate on the relationships of Nahali with other Asian language isolates (Kusunda of southern Nepal; Ainu of Sakhalin and Hokkaido of the Japanese archipelago; Andamanese; language of Vedas in Sri-Lanka), and waver between Nahali as proper isolate (van Driem, 2001), or as an Austroasiatic language (Mundlay, 1996), or rather as an argot spoken by only a minority of ethnic population (Zide and Barker, 1966). The occurrence of these linguistic outliers on the crossroad between the three major linguistic families spoken in South Asia (the Indo-European, the Dravidian, and the Austroasiatic) and along the margins of what has been proposed as a "shortcut" for modern human dispersals towards Southeast Asia in alternative to the coastal route (Field et al., 2007) highlight their importance in South Asia. Today, very few people speak the Nahali language and are mostly confined to the Jalgaon-Jamod area, at the bottom of the Southern slopes of the Satpura

range (Crivellaro, personal communication). One of the papers by the Anthropological Survey of India (Kumar et al., 2008) reports a 'Nihal' population with striking similarities with Kathodi and Katkari populations of Gujarat based on comparative analyses of complete mtDNA polymorphisms. It is not clear, though, whether the ethnonym 'Nihal' refers to the Nahali linguistic isolate or to another homonymous population affiliated to the Indo-European language family found in Nandurbar and Jalgaon districts of Maharashtra (Lewis, 2009) – see table 1 from Kumar et al., 2008). The genetic status as well as detailed grammatical study of Nahali still remains one of the highest research priorities among geneticists and linguists.

Burushaski is another language isolate spoken in the central Hunza valley of Northern Pakistan.Till date, there is no script available for Burushaski and original historical material is very scarce. Habitat separation from their neighbors in the Karakoram Mountains was the major reason which might have preserved their language with lots of borrowed words (van Driem, 2001). There are many myths behind the origin of these people (compiled in Mohyuddin et al., 2006). Concerning molecular genetic studies from these people, there are abundance of data available (Qamar et al., 2002; Quintana-Murci et al., 2004; Rosenberg et al., 2002; Rosenberg et al., 2006; Sengupta et al., 2006; Li et al., 2008). Incorporation of Burursho samples in HGDP panel was the most important step towards knowing the genetic makeup of this population. This is the reason, why Burushaski is the single genetic isolate of South Asia, whose genetic structure is much explored (Rosenberg et al., 2002; Rosenberg et al., 2006; Li et al., 2008; Auton et al., 2009; Ayub and Tyler-Smith, 2009). The recent high resolution autosomal study based on 650K SNP chip suggested their largely South Asian ancestry with moderate East Asian genetic input (Li et al., 2008). It was suggested that their genepool is heavily diluted by genetic exchange with their neighboring populations (Ayub and Tyler-Smith, 2009). This is also reflected in their vocabulary which has many words borrowed from their neighbors (van Driem, 2001; Blench, 2008). Nonetheless, it is important to note that HGDP has only 20 samples from this population and high resolution Y-chromosomal study reported 5% and 20% individuals belonging to haplogroup K* and R* respectively, which are however rare in South Asia (Sengupta et al., 2006).

Apart from the linguistic isolates described above, there are few intriguing populations who greatly differ from their neighboring populations. Bonda and Didayi are two examples residing in Malkangiri district of Orissa state. Bonda is also called as highlanders population living at Bonda-hill situated at the northwest of Machkund river. They are classified as a tribe of Austroasiatic (Munda) language family, speak Koraput-Munda branch of Southern Munda sub-family (Fuller, 2007) and spread across 32 villages in the Khairput block of the Malkangiri district. Bondas who live above the hill are known as upper Bonda and are still hunter gatherers, while the Bondas who subsist down the hill are called lower Bonda and perform jhuming cultivation. The number of females is more than the males and they also hold higher status in the

community. The dress of Bonda women is unique and remarkable. It should be noted that they have very peculiar marriage system where a young girl (aged around 20 years) marries a boy 8–10 years younger than herself. Didayi are also found in the same district on bank of Machhakunda river in the centre of Kanda-kamberu hills. They are mainly a hunter-fisher population speaking language similar to that of Bonda. Didayi is residing on the left bank of the river Machhkund on a small island which is cut-off from the mainland as a result of the Balimela water reservoir. The process of wine preparation is fascinating and similar in both the populations. Jackfruits and mangoes are subjected to fermentation for three consecutive days and then distilled over the natural small waterfalls. Women of both populations play an active role in bringing up their families. Extensive molecular genetic study is lacking in above tribes, however, a few studies do report lower frequency of G6-PD deficiency than their neighbors (Balgir, 2006,2008; Nishank et al., 2008) which is intriguing considering the fact that they are living in malarial endemic area since several thousand years, where the G6-PD deficient gene should have been advantageous. Apart from the aforementioned tribes, Mankiridia of Orissa, Malani of Kullu district of Himachal Pradesh, Abujhmaria from Bastar district of Chhatishgarh, Chola from Kerala states of India are equally important in having their unique entity in comparison to their contemporary populations (Fig. 5).

Figure 5. Geographical location of language isolates and other enigmatic South Asian populations.

Thus, it is evident from the above chapter that the linguistic separation of a population is not necessarily indicative of its genetic uniqueness. Because language changes so rapidly, and none of the presently spoken languages can be singled out or considered in any reasonable sense to be relevant to late Palaeothic population dispersal. In some cases, for example in case of Andaman islanders, the long term geographical separation and other genetic factors such as drift were the major factors to create unique variation. Several genetic studies have been done to test the concordance between gene, language and geography among South Asian populations (Kivisild et al., 2000b,2003a; Cordaux et al., 2004c; Sahoo et al., 2006), and gene and geography do show a significant correlation (Ref. IV).

2.3.6.1. Gypsies: A recent offshoot

Apart from enigmatic populations restricted to South Asia, there is another intriguing group which has extended itself from Middle East to Europe. A single divergent population (or ethnic group) harboring the vast geography of Europe (in their patchy distribution), who speak different languages and are never connected with any national identity are “Gypsies”. They are also known as Roma. The term "Gypsies" originated from the popular belief that they have Egyptian origins for which, however, there is no proof. Their culture is highly oriented to art and music with many mythical stories and songs. Sareen (2005) mentioned a nice quote about them that they “*never sleep twice in the same place never drink water twice from the same well, and never cross the same river twice in one year.”* The history and migration of this socioculturally diverse ethnic group is now getting clearer with the accumulation of new datasets coming from evolutionary as well as medical genetics (for a recent review see Kalaydjieva et al., 2005 and references therein). The European Gypsies are a recent founder population who have migrated “Out-of-India” more particularly from Northwestern region, probably one thousand to fifteen hundred YBP and reached Balkans through Middle East (Fraser, 1992; Rochow and Matschke, 1998), though their mythical story suggests that the Roma are from the Ganges belt of India. The genetic analysis of Euroepan Roma populations identified them as an intriguing founder population to study Mendelian disorder and a valuable part of the European genetic landscape (Mastana and Papiha, 1992; Kalaydjieva et al., 2001,2005; Mihaylova et al., 2007; Bouwer et al., 2007).

The Indian ancestry of Roma has been already extensively studied and attested in a number of mtDNA, Y-chromosomal and autosomal studies (Gresham et al., 2001; Zhivotovsky et al., 2004; Kalaydjieva et al., 2005; Malyarchuk et al., 2006,2008; Mihaylova et al., 2007; Bouwer et al., 2007). The presence of mtDNA haplogroups M5 and M35 (Malyarchuk et al., 2008), single Y-chromosomal haplogroup H1a (Gresham et al., 2001; Klarić et al., 2009) and the pathogenic 1267delG mutation in CHRNE (cholinergic receptor, nicotinic,

epsilon) (Abicht et al., 1999), are rooting their ancestry to South Asia. It is also evident that they have exchanged significant amount of genes from their contemporary populations on their way to Europe (Klarić et al., 2009 and references therein).

Thus, the molecular studies have helped us to understand the origin of Roma and the next step should be focused towards finding the most recent common ancestral (MRCA) South Asian population to them. South Asia also has many local Gypsy populations *viz.* Bawariya, Gujjar, Kanjars, Dharkars, Nut, Haburas, Bediyas, Baheliyas, Bhantus in the Ganga plains; Ghumantu, Sansis, Bawariya and Kalbeliyas in Rajasthan; Chharas in Gujarat; Pardhis and Kuchbandhias in Madhya Pradesh, Pardhis and Vaidus in Maharashtra; Lambadi in Andhra Pradesh; Kallars in Tamilnadu and Gujjars and Hurs in Pakistan. They are known as denotified and nomadic tribes, although before the independence they were known as criminal tribes (Ref. II). In India these populations are also identified as the most notorious tribes. They are commonly known as `*Kaccha-baniyan'* gang or `*Bawariya*' with their peculiar *'Modus operandi'* for theft and murders. A thorough genetic study encompassing molecular comparison between Roma and South Asian nomadic tribes should however be able to throw some light on the relation and recent split of these groups.

2.3.7. The Social/Caste System and its impact on genepool

South Asia constitutes an incredible large segmented society that harbors rich genetic diversity within its human populations and provides a fundamental platform to study the various factors influencing demographics. Strict endogamous populations, flourished independently with varied socio-cultural and linguistic diversities, nurtured by the vast geographical and ecological systems enriching South Asian population history tremendously. It is of particular interest to study patterns of genetic affinities among various endogamous groups inhabiting small geographical regions within the subcontinent because of their diverse origins and interethnic separation. The strict practice of endogamy across all social ranks resulted in emergence of population-specific diverse social traditions and development of distinct linguistic dialects. These divergent socially structured population groups provide a varied substratum for understanding variation of a genetic trait, spread of a particular disease and their prehistorical settlement in a geographical area resulting from interaction among various population groups. In general, human beings group themselves into units in such a way that members between units rarely exchange genes due to cultural and geographical barriers resulting in genetic divergence of populations. This phenomenon is termed as Endogamy. Therefore, one can explain endogamy as a phenomenon in which populations do not share genes even though they live at the same social level and exchange the rituals and other traditional culture with one another and maintain their unique social/genetic identity. Endogamy is practiced by nearly all communities worldwide but the

South Asian endogamy is unique in itself by harboring several layers of endogamous barriers. The simplified structure of social boundaries and marriage system of South Asian groups has been explained in figure 6.

South Asia is traditionally known as the classical land of castes. Caste is a social institution, deriving sanction from, and intimately interwoven largely with the Hindu religion. The Caste sanction and structure govern all social, religious and economic activities of people. The Caste stratification since the RigVedic society has been based on the *Chaturvarna* doctrine *i.e.* Brahmin, Kshatriya, Vaishya and Sudra. Brahmins are the torch-bearers of religion and perform rituals, and are involved in learning and teaching of the society. Kshatriya ranked next as rulers, defenders and warriors, Vaishya became cultivators, artisans and traders. These first three castes are called as *'dwij'* (twice born) *i.e.* after attaining an age (after *upanayana sanskara,* people receive the sacred thread '*janeo*') they are allowed to learn Veda, where they are considered to be reborn. The fourth class is known as Sudra, who works as labors in the society. Information about early social organization is composed in the form of Veda which is considered as oldest scripture of Hinduism. *Rig Veda* is believed to have been composed circa 1700–1100 B.C. (Rao et al., 2009). These hymns are represented in Devanagari script. However, the chronology of Veda is not in general acceptance and has been always a matter of dispute (Wheeler, 1979; Possehl, 1979; Kazanas, 1999; Talageri, 2000; Lal, 1997,2009; Witzel, 2005). It is argued that the Vedic literature had grown in the course of several thousand years and was handed down from generation to generation by word of mouth (Banerjee, 1979; Glacier, 2006), and due to the short life manuscript materials which were made up of birch bark or palm leaves usually, surviving manuscripts rarely surpassed an age of a few hundred years (Brodd and Gregory, 2003). Apart from Vedic literature, there are Puranas and Upnishads which give us clue about such structuring. The *RigVeda* broadly deals with ritualistic (elaborated in *Brahamanas*) and philosophical aspect (elaborated in *Upnishads*). Another debate is about the present structure of the caste system. Some of the scholars suggest that the Vedic classification of caste was based on the occupation of the individuals referring to hymn (9:112) of *Rigveda* (Thapar, 1966; Carvalho-Silva and Tyler-Smith, 2008). Alternatively, this hymn can be explained on the philosophy of *"Vasudhaiva Kutumbakam" i.e.* the theory of one world (Hatcher 1994; Saraswati and Bhavan, 1995; Pratap, 2001).

The caste system in South Asia is the topic which has been extensively studied by several disciplines of humanity, but it is also true that several assumptions made it one of the most controversial cases. The term *caste* comes from the Portuguese word *casta* (breed, lineage) and was coined by Portuguese after witnessing the social, economic, and religious system of South Asia. The traditional Hindu term is *varna,* and its meaning is *to enclose* or *color*. In each and every geographical region of South Asia, there are about hundreds of groups called castes with distinct names, birth in one of which usually determines the status of a given individual in society. Each caste is further divided into smaller units – generally known as sub-castes and finally these are then divided into multiple Gotras (purely exogamous clans) that effectively prohibit

consanguineous marriages. It is an exogamous character of a clan that makes it a distinctive group (Fig. 6). In practice a Gotra consists of a large number of *pravaras* supposed to be descended from the same sage-ancestor, who lived in the ancient past. In time the number of decendants of each sage-ancestor had increased which affected all ties of common residence and even territorial proximity was snapped, but a large number of new derivative Gotra kin groups have been renamed after new sages. It is not merely the relationship that determines the association into a Gotra, but there are few exceptions where the disciples of a sage adopted the same Gotra as that of their preceptor. The Gotra system determines which clans can marry with each other and facilitates thereby effective social life and substructure within the caste. As long as Gotra was regarded as a group of consanguineous kin, the question of marriage within the Gotra (*Sagotri* marriage) didn't arise (Fig. 6). The Gotras are commonly related with specific cultural traditions. For example, Brahmins trace their ancestry to one of seven sages believed to have been connected to the Vedic system (Glacier, 2006). Knowledge of the caste and Gotra is in fact essential to perform any rituals in the society.

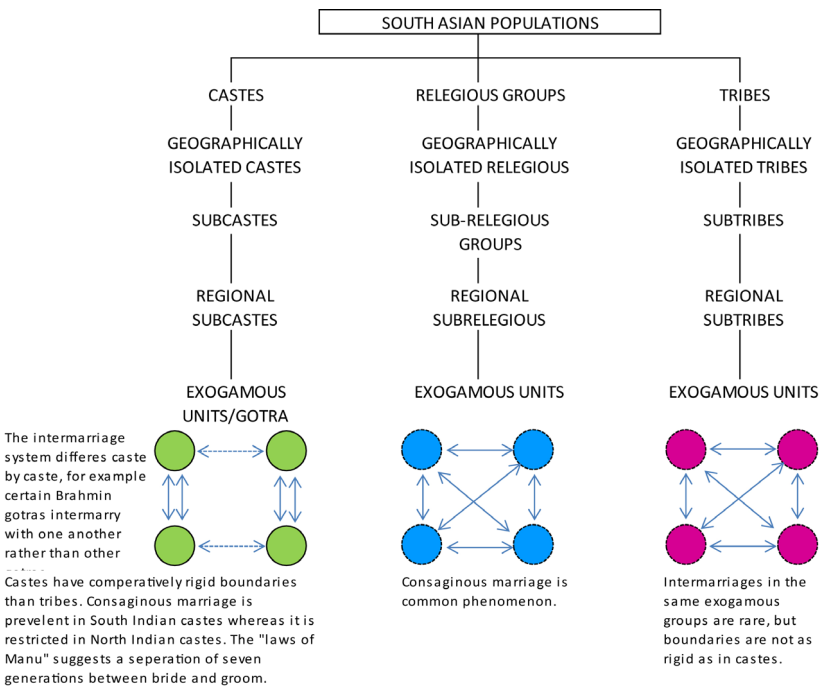

Figure 6. The simplified representation of South Asian population's social structuring. Structure drawn based upon the society definition given in various Hindu scriptures published by (Gita Press, Gorakhpur India; Vivekananda, 1970). Each strict exogamous unit and movement of spouces are shown by solid circles and bidirectional lines, respectively, while, the dotted circles and bidirectional lines depict the permeable boundaries within the exogamous unit.

The central dogma of Vedic caste system has been explained in various Hindu scriptures by numerous Hindu scholars (Gita Press, Gorakhpur India; Ashrama, 1944; Vivekananda, 1970; Kutty, 2006). These scriptures suggest that the caste system was made to raise the level of lower caste and make everybody Brahmin (here Brahmin doesn't mean as a name of a caste, it designates the ideal person of humanity) (Vivekananda, 1970). The Hindu mythological sources mainly pinpoint on the immortality of soul which experience maximum 96000 births (Gita Press, Gorakhpur India; Ashrama, 1944; Vivekananda, 1947) (Purana, 1961; Goyandka, 1969; Vivekananda, 1970). These sources further explain that each and every person born in any caste is the messenger of a particular soul and one's birth in the present life is a direct consequence of knowledge gained and *Karma* accrued from their previous lives. If a person performs bad *Karma,* he/she will be born in the lower caste but if he/she does a better Karma he/she will be born in an upper caste. Thus, caste system inspired an individual to perform a better karma.

Caste system in South Asia fabricates cultural barriers and segregates populations into several endogamous pockets which restrict the movement of spouses, thus, limiting the gene flow outside the caste. There are several genetic studies related with South Asian caste system (Bamshad et al., 2001; Basu et al., 2003; Kivisild et al., 2003a; Metspalu et al., 2004; Cordaux et al., 2004a,c; Sahoo et al., 2006; Sengupta et al., 2006; Thanseem et al., 2006; Ref. II and III). A recent series of papers based on mtDNA and Y-chromosomal analysis instigate that the origin of the caste system was mainly rooted to male-mediated migration (Indo-Aryan invasion) who pushed indigenous Dravidian speaking populations towards Southern India and Sri-Lanka, and established Indo-Aryans as upper caste people (Wells et al., 2001; Quintana-Murci et al., 2001; Bamshad et al., 2001; Basu et al., 2003; Cordaux et al., 2004a,c; Sahoo et al., 2006; Sengupta et al., 2006; Thanseem et al., 2006). It was added that the caste populations are closer to Europeans and Central Asians and significantly differ from tribal populations (Bamshad et al., 1998,2001; Cordaux et al., 2004c). However, these papers were tainted with low resolution and limited sampling size, but succeeded to initiate a trend for population geneticists to investigate the integrity of caste system. Most of the researchers, either they have supporting datasets or not, started including the Indo-Aryan invasion concept in their manuscripts, assuming that it as a universal historical proven fact (Quintana-Murci et al., 2004; Sharma et al., 2005; Thanseem et al., 2006; Zerjal et al., 2007; Ayub and Tyler-Smith, 2009; Chakravarti, 2009). A recent review summarizes briefly the different caveats of study of caste system in South Asia (Boivin, 2007). However, there is now universal concurrence that various castes and tribal populations of India share a common late Pleistocene maternal as well as paternal ancestry (Kivisild et al., 2003a; Metspalu et al., 2004; Sahoo et al., 2006; Ref. I–IV, which was improvised by minor geneflow from East and West (Metspalu et al., 2004; Ref. II). Another important aspect of the Indian social system, which makes substantial impact on the inferences one can make from the caste/tribal genetic variation, is that the definition of the scheduled caste has

not been very clear. Specifically, the scheduled tribes were gradually incorporated into the caste system as scheduled castes. The social uplifting process thus enhanced a gradual absorption of tribal populations into the caste system (Ref. II and III). Above studies suggest that caste system which divide populations into several endogamous pockets, has undoubtedly played an essential role in shaping South Asian genetic architecture.

2.3.7.1. Family as a functional unit of Society

Family is a common word used in any human society and in simple sense it is well understood. Family has a biological matrix. It is the smallest functional unit of a society which includes all the people living under the same roof. The family is based on simple facts that involve the recognition of those who are closely related with one another by blood ties. More elaborately, it refers to a universal, permanent and persistent institution characterized by socially approved sexual mating and reproduction, common residence and economic cooperation. Every kind of human society survives and is maintained through a process of recruitment, with the help of which various individuals become its members in a cyclic process. Majority of members become an inherent part of family by being born into the family while others get an entry through the process of marriage (*e.g.* females in patrilocal societies and males in matrilocal societies). There is a great diversity in the forms of family in South Asian society. These forms differ from each other on several different bases *viz.* residence, descent, membership and mates. In most communities in South Asia descent is traced along the paternal line which is also practiced worldwide. Though, matrilocal system does also exist in some populations *e.g.* Garo, Khasi tribes of Northeastern states, Nayars in Kerala and Lakshadweep islanders. In a number of tribal groups, similar to Western societies, a neolocal residence practice (newly-wedded couple move to new house) is preferred. In the definition of family, 'joint family' term is often used, particularly in rural India which is by and large absent in Western world. It is a composition of two, three or more generations of lineally related members. In that case several nuclear families (a married couple and their children) live together and have common arrangements for cooking and dining.

Family types may also be classified on the basis of number of mates. Most of the societies follow monogamy in which an individual can have only one spouse at a time. While polygamy describe having more than one spouses. It can again be divided in to two sub categories: polyandry, where a woman can have two or more husbands, and polygyny, in which a husband can have two or more wives. Muslim personal laws permits polygyny in India while Pakistan has passed legislation intended to eliminate the abuse of polygyny. Polyandry is rare, however, it is present in some small communities, for example, Jaunsaris of Uttranchal and Kannaur-Lahaul in Himanchal Pradesh. Among them, when

the eldest brother marries a woman, all brothers younger to him simultaneously become her husband.

Where family and marriage comes in discussion, the choice of mate is another intriguing topic. However, there is not much information available about it in the early modern humans. Genetic studies suggest small founding groups for human populations (Kruglyak, 1999; Macaulay et al., 2005), therefore, significant inbreeding can't be overlooked (Bittles and Black, 2010). In mid 19th-century, marriage among cousins was socially accepted and widely favored (especially among the more privileged classes), for the first time in Europe and North America (Ottenheimer, 1996; Bittles, 2003). In recent times, however, consanguineous (cross-cousin) marriages are recorded in several parts of the world (for a worldwide view see http://www.consang.net/index.php/Global_prevalence). It is evident that the highest rate of consanguineous marriage occurs in North and Sub-Saharan Africa, the Middle East, West, Central, and South Asia and it has been reported as 10.4% worldwide (Bittles and Black, 2010). In South Asia, the Vedic culture of Hindus allow marriage only among people who are seven generations apart and belong to different Gotras, while in Southern Indian Hindu societies, marriage between close relatives (cross-cousin) is permitted in several populations. Consanguinity therefore, is not restricted only to Islamic communities.

The influence of cultural practices shaping the genetic pattern of the society has also been tested by several molecular studies (Seielstad et al., 1998; Rosser et al., 2000; Oota et al., 2001; Kayser et al., 2001a; Kumar et al., 2006a). It was estimated that approximately 70% of modern societies practice patrilocality (Seielstad et al., 1998). Recent studies illustrate a high degree of inter population genetic difference for Y-chromosome compared to mtDNA due to patrilocality practices (Seielstad et al., 1998; Oota et al., 2001). However, sex based differential migration may not be a significant factor shaping the observed global distributions as proposed by Seielstad et al. (1998). This phenomenon has been used in interpreting the clinal pattern of Y-chromosomal variation in Europe (Rosser et al., 2000) and in islands of Southeast Asia (Kayser et al., 2001a). There are also some examples where mass migration of male lineages changed the complete scenario of paternal genepool, for instance, the expansion of Europeans into Americas and Oceania in the last 400 or more years adding European Y-chromosomes with retention of indigenous mtDNA lineages that is seen in Polynesia (Hurles et al., 1998), Greenland (Bosch et al., 2003) and South America (Carvajal-Carmona LG et al., 2000; Carvalho-Silva et al., 2001). Oota et al. (2001) suggested that variation in mtDNA and Y-chromosome are different in matrilocal and patrilocal communities in Thailand. In contrast to this, recent study on Indian populations by Kumar et al. (2006a), found no correlation between patrilocality and matrilocality in Indian populations. They suggested that a sex specific difference in migration rate is not a primary factor influencing the discrepant patterns of genetic differentiation.

Interlocus comparisons assume that migration rate is a major factor determining population differentiation, however, random genetic drift or founder

effect also influence the level of population differentiation (Ref. II). The consequences of genetic drift are especially accentuated in those social situations where any chance of some degree of random mating are extremely minimal, and therefore dictates that, ones who have allowed social mating has privilege while the who didn't, are rigidly abandoned. Hence, it was suggested that a sex specific difference in migration rate is not primary factor influencing the discrepant patterns of genetic differentiation. In India, the marriage practices are restricted within an endogamous population consisting of several exogamous clans or Gotras (see Fig. 5). And the marriages take place between the hierarchical orders of these Gotras. Therefore, it is not applicable in Indian panoarma where marital boundary is highly impermeable. Besides this, gene-culture interactions had a prominent role in local, geographically restricted adaptation over the past several thousand years (Ref II–IV).

2.4. Molecular and evolutionary Genetics on origin of modern humans

Every human being carries within each of their cells, a long encoded history of the species and their ancestor(s). The human genome holds an online record of evolution in itself that stretches back to the first *Homo sapiens* and beyond, to the earliest primates, the first animals, and the origin of life itself. Understanding human genome not only requires comparing two individuals (populations), but also comparing them with other species. Organisms have inherited most of their genes and body building processes from their common ancestors. Some of the genes (autosomal) undergo reshuffling while some of them (*e.g.* mtDNA and NRY) pass without any recombination. This chapter briefly introduces the role of haploid markers in uncovering the Out-of-Africa migration and evaluates the question of inbreeding between modern humans and Neanderthals. Subsequently, in the forthcoming chapters the pre-DNA and DNA-era of modern human dispersal in South Asia will be discussed concluding with a final review on high resolution autosomal studies.

2.4.1. Out-of-Africa replacement and modern human dispersal

The recent single African origin of modern human is no more a point of discussion and considerably supported by all the disciplines of human studies, *e.g.* archaeology and fossil evidence (Stringer, 2000; White et al., 2003; Mellars, 2006a,b), ancient DNA (Adcock et al., 2001; Hofreiter et al., 2001) and modern DNA/genetics (Cann et al., 1987; Tishkoff et al., 1996; Ingman et al., 2000; Kivisild et al., 2003a; Metspalu et al., 2004; Thangaraj et al., 2005a; Underhill et al., 2000; Underhill and Kivisild, 2007; Macaulay et al., 2005; Ref. I), with the exception of few supporting multiregional model (Eswaran et al., 2005; Harpending and Eswaran, 2005). It is well accepted that all non-African popu-

lations were derived from a recent colonization (~50–70 KYA) of anatomically modern humans (AMH) from Africa. It has been suggested that all non-African populations contain only a subset of genetic diversity present in Africa (Tishkoff et al., 1996; Behar et al., 2008a). In other words, all the non-African maternal lineages (M and N) are derived from a single branch (L3) exclusively present in Africa (Mishmar et al., 2003; Kivisild et al., 2004; Macaulay et al., 2005; Thangaraj et al., 2005a; Behar et al., 2008a). Similarly, the pan-Eurasian paternal lineages are also derived from an African branch carrying CT-M168 polymorphism (Underhill et al., 2000; Jobling and Tyler-Smith, 2003; Underhill and Kivisild, 2007). After the wide acceptance of Out-of-Africa migration theory the debate shifted towards the number of dispersal events which occurred (Lahr and Foley, 1994,1998; see also a review by Richards et al., 2006). Two dispersal events were suggested: one with Southern route from the horn of Africa and another through Levantine corridor (Lahr and Foley, 1994) which recently was bolstered by an archaeological discovery, dealing with Toba ashes in South Asia (Petraglia et al., 2007). Yet, mtDNA and Y-chromosomal studies are in support of a one wave exit of modern humans along the Southern route (Metspalu et al., 2004,2006; Macaulay et al., 2005; Forster and Matsumura, 2005; Palanichamy et al., 2004; Sun et al., 2006; Hudjashov et al., 2007; Chandrasekar et al., 2009). Similar scenario has also been supported by archaeological studies (Mellars, 2006a,b). More importantly, it was also observed that the indigenous lineages reported in South Asia, Southeast Asia, East Asia and Oceania directly pop-out (star like phylogeny) from the root of the pan-Eurasian founders (M and N), tempting scientists to propose "Express-Train" model (Macaulay et al., 2005).

From their birthplace in Africa, modern humans spread to Asia, Europe, and rest of the world. There are two hypotheses connected with range of expansion of AMH: either new groups arriving from Africa "absorbed" older hominid branches through interbreeding (Nordborg, 1998; Templeton, 2002,2005; Mellars, 2006a) or global replacement and genetic extinction of non-modern human populations by AMH (Stringer and Andrews, 1988; Fagundes et al., 2008). The discovery of Neanderthal and successful extraction of its DNA was a major step towards assessment of the above hypotheses.

2.4.1.1. Neanderthal mtDNA genome and its relation with modern human

Neanderthal is an extinct member of the genus *Homo* recognized as a distinct group of hominids from fossil remains for the first time in Germany (Neandertal Valley near Düsseldorf) about 150 YBP (Green et al., 2006 and references therein). So far, the fossil of Neanderthals has been unearthed from Europe and parts of Western and Central Asia (Krause et al. 2007). By 130,000 YBP, complete Neanderthal characteristics had appeared. These characteristics then disappeared in Asia by 50,000 YBP and in Europe by 30,000 YBP (Green et al.

2006). This time overlaps with the existence of modern humans in West Asia and Europe (Krings et al. 1999; Mellars 2004). The sequencing of Neanderthal mtDNA allowed molecular testing of many hypotheses concerning the intermixing between modern humans and Neanderthals (Nordborg, 1998; Templeton, 2005; Mellars, 2006a and references therein) as well as the Out-of-Africa replacement model of modern human origin (Stringer and Andrews, 1988). The time of origin of modern humans as a species and their relationship with other hominin species are some of the most basic questions in the field of human evolution. The discovery of Neanderthals and extraction of its mtDNA has been one of the most significant contributions towards solving these questions. Nevertheless, it should be noted that ancient DNA studies had major pitfalls in many of the previous works (Gilbert et al., 2005), and only until recently it has started gaining support after the arrival of high thoroughput sequencing technologies (Prüfer et al. 2010; for a review see Ho and Gilbert, 2010).

The first step taken to understand the parallel existence or evolution of Neanderthals and modern humans was sequencing of Neanderthal mtDNA HVS-I region (Krings et al., 1997). Furthermore, similar mtDNA studies from other Neanderthal specimens were also done (Krings et al., 1999; Serre et al., 2004). The comparison of this segment with modern human mtDNA didn't reveal any introgression between these two hominid species (Serre et al., 2004). The statistical analysis based upon coalescent models also supported the lack of introgression of Neanderthal with modern human (Currat and Excoffier, 2004; Blum and Rosenberg, 2007). With the technological improvement, some more specimens of Neanderthals were sequenced for the HVS-I region which suggested that all Neanderthals form a monophylectic clade, and were separated from the modern humans several thousand YBP (Ovchinnikov et al., 2000; Krings et al., 2000; Lalueza-Fox et al., 2005; Green et al., 2006).

The use of merely HVS-I information was not sufficient to resolve the questions concerning genetic relation between Neanderthals and modern humans. As it is evident in case of modern DNA, the smaller size and greater variation of HVS-I can't provide sufficient information to distinguish many important ancient branches within the tree. Moreover, the high level of variation in the HVS-I region along with numerous recurrent mutations can distort the structure of a phylogenetic tree. For example, many deep rooting mtDNA lineages (M31,M33,M35,M43,R8,R30,R31 *etc.*) in South Asia have entirely different HVS-I mutations at different branches (Palanichamy et al., 2004; Thangaraj et al., 2005a,2008,2009; Palanichamy et al., 2006; Sun et al., 2006; Chandrasekar et al., 2009; Ref. I,II,IV). Thus, HVS-I information solely is certainly not the key point to supply adequate information to rule out if the admixture occurred 30,000 YBP or not (Nordborg, 1998). Furtheremore, after the death of an animal, DNA starts to decay rapidly resulting in reduction of number and quality of surviving DNA fragments, thus, increasing rates of sequencing error and forming a significant obstacle to accurate sequence reconstruction (Gilbert et al., 2005; Ho and Gilbert, 2010). To rule out such errors, DNA from contem-

porary early modern human fossil and Neanderthal DNA were amplified and sequenced with Neanderthal specific primers (Serre et al., 2004). The results obtained were consistent with the previous studies, thus ruling out any possibility of Neanderthal-modern human admixture during prehistoric times.

Recently, two groups independently analyzed nuclear genome of a 38KYA old Neanderthal (Green et al., 2006; Noonan et al., 2006) and landed to different conclusions. Green and his colleagues (Green et al., 2006) obtained one million base pairs of Neanderthal DNA and calculated divergence time with modern humans as 516KYA. The comparison with HapMap data surprisingly led to 30% SNP similarity with modern humans, which was a major evidence of admixture between modern human and Neanderthals. In contrast (Noonan et al., 2006) obtained ~65Kbp genomic sequence and estimated the divergence time to 716KYA. The comparison with HapMap data here showed only three derived human SNP variants, two of which were found in African populations. This resulted in absence of admixture between modern humans and Neanderthals. The reanalysis of (Green et al., 2006) data later with maximum likelihood approach revealed a modern human DNA contamination as high as 78% (Wall and Kim, 2007). A more recent study sequenced >4 billion nucleotides from three Neanderthal individuals along with five modern human samples from different continents and identified several positively selected regions (Green et al. 2010). This study also suggested a gene flow from Neanderthals to the ancestors of non-Africans before the divergence of modern human ancestors of Eurasian decendents (Green et al. 2010). However, more research is needed from modern and ancient DNA to clarify this debate.

2.4.2. From classical to haploid markers: Fact file from South Asia

2.4.2.1. Pre-DNA era

The A, B and O blood groups, discovered by Carl Landsteiner in 1801, were first used to study genetic variation in modern humans. The reconstruction of human prehistory by the use of allelic distribution was first attempted by use of gene frequencies of 20 alleles from five major blood-group systems known from 15 populations (Cavalli-Sforza and Edwards, 1967). However, later with the advancement of technology, protein and enzyme polymorphisms were discovered and used for genetic tree reconstruction (Nei and Roychoudhury, 1982). An extensive work on classical genetic polymorphisms (more than 100 genes) on South Asians in global perspective placed South Asian populations between the populations from Southeast and West Asia (Cavalli-Sforza et al., 1994). This study included Dravidian and Indo-European speakers of the Indian subcontinent. There are two major studies from classical era of South Asia which can be highlighted here-one by ICMR (Indian council of Medical Research), is a compilation of classical data from 191 major tribal populations

(Bhatia and Rao, 1986) which covered several polymorphic traits (Blood groups, Red Cell enzymes, Serum protein and diseases *viz.* Hemoglobin disorder and G6-PD deficiency), another summarized by S.S. Papiha (Papiha, 1996) which suggested a distinct differentiation between caste and tribal populations. All India Anthropometric Survey (with 13 measurements) has also been published by Anthropological Survey of India (ASI) (Sreenath and Ahmad, 1989).

2.4.2.2. DNA era

An understanding of evolutionary history of peopling of South Asia has long been a subject of interest for the researchers working in evolutionary and medical genetics. The relation amongst different language groups and their origin is one of the most highly debated issues. But it should be taken into consideration that none of the language families in the world is as old as mtDNA or Y-chromosomal age of any haplogroup and therefore should not be related with linguistic affiliations. Based on results of haploid genetic markers *e.g.* mtDNA and Y-chromosome, it is widely accepted that the South Asian genepool is largely autochthonous (Ref I-IV and references therein). Genetic analyses of classical and haploid markers provide opportunity to examine population history and genetic structure (*e.g.* Cavalli-Sforza et al., 1988; Kivisild et al., 2006b; Underhill and Kivisild, 2007). Population level at first and individual level variation now is gaining a common place for association studies and analysis of human evolutionary histories. The beginning of DNA-era was a major breakthrough to compare principally this segment in different individuals. The underlying principle of this approach is to reconstruct the history of changes (mutations) found in the DNA of contemporary individuals, and to trace their origins to a most recent common ancestor(s) (MRCA) who would have lived at some point in the past.

The DNA-era in South Asia began with the study of few RFLP markers and their comparison with other Asian, Caucasian and African populations (Harihara et al., 1988; Semino et al., 1991; Soodyall and Jenkins, 1992), and further elaborated with the advent of PCR and sequencing techniques (Mountain et al., 1995; Bamshad et al., 1996,1998; Passarino et al., 1996). The preliminary studies were unable to give a clear picture regarding peopling of South Asia but the complexity of their genepool was evident. South Asia offered a canvas of genetic landscape to population geneticists, to test many unsolved complex theories amongst which the origin of caste and tribal populations and Aryan-invasion theory were the topmost priorities. Nonetheless, many other topics were also elaborated *e.g.* 9bp indel polymorphisms (Watkins et al., 1999; Clark et al., 2000; Thangaraj et al., 2005a; Kumar et al., 2006b), matrilocal-patrilocal communities (Kumar et al., 2006b), agricultural expansion and language shift (Cordaux et al., 2004a; Ref. II), origin of Andaman and Nicobar islanders (Prasad et al., 2001; Endicott et al., 2003,2006; Thangaraj et al., 2005a,b,

2006a,b; Barik et al., 2008; Reich et al., 2009), Y alu polymorphism (YAP) (Chandrasekar et al., 2007), phylogeny based on complete mtDNA sequencing (Palanichamy et al., 2004; Sun et al., 2006; Chandrasekar et al., 2009; Ref. I), detailed study of a particular haplogroup (Kumar et al., 2008; Thangaraj et al., 2009; Underhill et al., 2010; Ref. IV) and origin of other minorities such as Muslims, Siddhi, and Jews communities (Thangaraj et al., 1999; Gutala et al., 2006; Sengupta et al., 2006; Terreros et al., 2007; Behar et al., 2008b;2010; Eaaswarkhanth et al., 2010). In the beginning of 21st century, main discussion about South Asian population was centered around two school of thoughts dealing with the origin of Indian caste and tribal populations. The first model suggested that the South Asian tribal and caste populations share significant Pleistocene heritage (Kivisild et al., 2003a), with limited recent gene flow while second one suggested their independent origin (Cordaux et al., 2004c). Recent advancement in data analysis methods, complete mtDNA sequencing and discovery of new Y chromsosmal SNPs, greatly increased the accuracy and resolution of population genetics analysis to test such models.

South Asian geography and contemporary populations make it an ideal place to study population dispersal event(s), testifying it as an extremely important place in the course of Out-of-Africa migration. Recently, it was suggested that most of the human populations lived in South Asia between 45-20KYA (Atkinson et al., 2008), which is also supported by archaeological discovery (Petraglia et al., 2009). In the past few years, genetic studies of mtDNA, Y-chromosomal and autosomal loci variation in South Asian contemporary populations have provided a substantial contribution to the understanding of human origin and dispersal patterns. Genetic studies on variation of mtDNA and Y-chromosome among human populations of South Asia provide evidence for a recent African origin and Southern route dispersal along the rim of Indian Ocean (Kivisild et al., 2003a; Endicott et al., 2003,2007; Metspalu et al., 2004,2006; Sun et al., 2006; Hudjashov et al., 2007; Chandrasekar et al., 2009; Ref. I). The current knowledge manifests that phylogeography of mtDNA macro-haplogroup (or super-haplogroup) M and Y-chromosomal haplogroup C provide an important piece of evidence for the Out-of-Africa migration of modern humans taking a Southern route to Australia (Kivisild et al., 2003a). Y-chromosomal haplogroup C consists of several sub-lineages with uneven phylo-geographic patterning, ranging from Central and North Asia to America and in the direction of South, Southeast Asia up to Oceania (Kayser et al., 2001a,b; Underhill et al., 2001; Sengupta et al., 2006; Hudjashov et al., 2007). Similar to mtDNA super-haplogroup M, this haplogroup has likely arisen somewhere in Asia after an early departure of modern humans from Africa and represents the signature of OOA migration (Sengupta et al., 2006). Haplogroup M is present from South Asia to Oceania, except its single branch *i.e.* M1, which is suggested to have a back migration from Asia (Olivieri et al., 2006). All the mtDNA lineages present between South Asia and Oceania are directly rooted to the two non-African founding lineages [M and N (R)] (Palanichamy et al., 2004; Thangaraj et al., 2005a; Macaulay et al., 2005; Merriwether et al., 2005;

Sun et al., 2006; Friedlaender et al., 2007; Hudjashov et al., 2007; Atkinson et al., 2008; Chandrasekar et al., 2009; Ref. I–II). Recently a nested structure was observed between Oceanian M42 and South Asian mtDNA sequences (Kumar et al., 2009). This exciting result implies that the migration from South Asia to Australia was not as rapid as proposed (Macaulay et al., 2005), or it could be possible that there was a second small scale migration subsequently later from South Asia to Oceania after pioneer settlement. However, it is quite unlikely and doesn't have any other support but more data from South Asia and Oceania would be able to throw some light on this recent enigma.

2.4.3. Dissecting the South Asian genepool applying modern genetic tools

Among South Asian countries, India and Pakistan are well represented in terms of genetic studies, however genetic data from Nepal, Bangladesh, Bhutan, Sri-Lanka and the Maldives are scarce. The genepool of Pakistani populations share a significant historical and pre-historical proportion of variation with West Asians and Europeans along with South Asian lineages (Quintana-Murci et al., 2004; Sengupta et al., 2006; Ayub and Tyler-Smith, 2009), while, Bangladesh shows similar kind of genetic composition due to its historical influence (Cordaux et al., 2003; Ferdous et al., 2009; Alam et al., 2010). Few studies on Nepal reported a deeply rooted ancestry from East with East and Southeast Asians as well as from West with South Asians (Gayden et al., 2007; Parkin et al., 2007; Kraaijenbrink et al., 2007a; Thangaraj et al., 2008; Fornarino et al., 2009). Except for a couple of forensic studies, Bhutan is also under represented for population genetic studies (Parkin et al., 2006; Kraaijenbrink et al., 2007b). Nonetheless, the complex geography of Himalayan states *i.e.* Nepal and Bhutan would require a further high resolution genetic dissection. Taken into consideration it's geographical position along the Indian Ocean as well as one and only fossil evidence provider of modern human in South Asia ((Kennedy, 2000), Sri-Lanka is equally important to study prehistorical human settlement, however, only few populations have been studied so far suggesting their migration from mainland India (Metspalu et al., 2004). Another island state *i.e.* the Maldives, has not been surveyed so far but the preliminary studies suggest their link with mainland Indian populations with high frequency of mtDNA haplogroup R30.

Over the last decade mtDNA and Y-chromosomal studies in South Asia have explored the affinity of populations within the large boundary of subcontinent as well as their relationships with other world populations (Kivisild et al., 1999,2003a; Quintana-Murci et al., 2004; Palanichamy et al., 2004; Metspalu et al., 2004; Thangaraj et al., 2005a,b; Sun et al., 2006; Sahoo et al., 2006; Sengupta et al., 2006; Atkinson et al., 2008; Chandrasekar et al., 2009; Ref. I–IV). Initial studies with adequate molecular resolution and sample sizes in South Asia suggested Austroasiatic speakers as initial settlers of the subcontinent (Majumder, 2001a; Basu et al., 2003; Kumar and Reddy, 2003), and further expanded towards

two rival hypotheses about genetic origin of caste and tribal populations, with more emphasis later on Aryan invasion (Kivisild et al., 2003a; Cordaux et al., 2004c). However, it should be noted that the present language families of South Asia (Indo-European, Dravidian, Sino-Tibetan and Austroasiatic), are all much younger than the majority of indigenous mtDNA and Y-chromosome lineages found at high frequencies (Ref. I-IV). Particularly, in case of India, there is no major distribution of mtDNA or Y-chromosome linguistically except for paternal lineages (haplogroup O2a) of Austroasiatic speakers (see in later chapter), though in Nepal the genes correlate well with language rather than geography (Yngvadottir, 2007). Similar studies with smaller number of populations, low power of genetic markers and without considering sociocultural factors provided the base for the delusion of molecular data without any solid base (Barnabas et al., 1996; Barnabas et al., 2006; Bamshad et al., 1996,2001; Basu et al., 2003; Cordaux et al., 2003,2004a,c; Baig et al., 2004; Gaikwad et al., 2005; Thanseem et al., 2006). For a critical postmortem of such studies see Boivin (2007) and Endicott et al. (2007). The advancement of technology (high throughput sequencing techniques), identified several deep rooting autochthonous lineages and helped to resolve several conflicting throries (Palanichamy et al., 2004; Metspalu et al., 2004; Sun et al., 2006; Sahoo et al., 2006; Sengupta et al., 2006; Chandrasekar et al., 2009; Ref. I–IV).

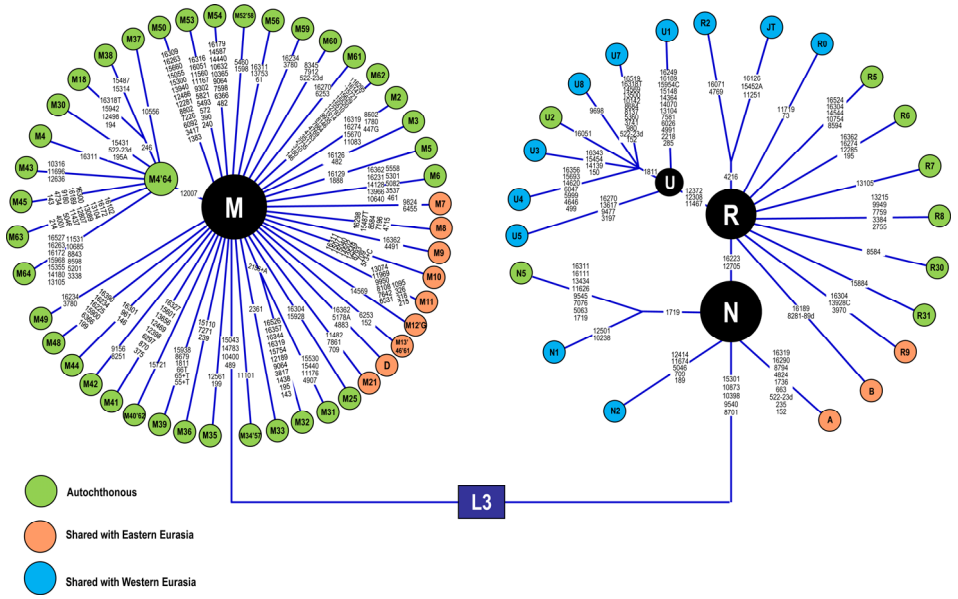

Figure 7. The deep rooting branching of autochthonous lineages, specific to South Asia and other lineages shared with West Eurasians and East Asians. Data obtained from Palanichamy et al. 2004; Sun et al. 2006; Chandrasekar et al. 2009; Ref. I–IV). The autochthonous lineages emerging directly from the root of pan-Eurasian founder super-haplogroups (M and N (R)) implying the fundamental role played by South Asia during intial OOA migration as well as testifying the autochthonous diversification of the maternal gene pool.

The maternal genepool of South Asia is largely composed of several indigenous deeply rooted lineages emerging directly from the basal node of two non-African founder lineages *i.e.* macrohaplogroups M and N, whereas R and U are nested in N (Fig. 6). A single 'Southern express route' was suggested from Sub-Sahara Africa along with Indian Ocean coast in to Eurasia and Oceania and later back migration to North Africa (Metspalu et al., 2004; Thangaraj et al., 2005a; Macaulay et al., 2005; Friedlaender et al., 2005; Olivieri et al., 2006; Hudjashov et al., 2007). Within South Asia some of the mtDNA haplogroups are concentrated to particular geographical region whereas, certain lineages are scattered all over the subcontinent (Metspalu et al., 2004; Chandrasekar et al., 2009; Ref. I–IV). East and West Eurasian lineages are also partially integrated with South Asian local genepool, however, they are concentrated alongwith the periphery of geographical boundary and share both ancient and young colescent age (Fig. 7) (Kivisild et al., 1999a; Metspalu et al., 2004; Thangaraj et al., 2008). For the first time a thorough study, based on complete mtDNA sequencing approach covering the large geography of Indian subcontinent (Palanichamy et al., 2004), revealed several indigenous deep rooting clades emerging directly from the root of superhaplogroup N and R. Several clades, such as N5, R5-R8, R30, R31 and U2a, were identified as indigenous to the subcontinent (Fig. 7). These large autochthonous haplogroups together with West Eurasian, Central and East Asian specific haplogroups present at low frequencies also suggests that the indigenous populations received only limited external gene flow from these areas (Fig. 7). Numerous studies on Indian populations emphasized on deeply rooted several mtDNA lineages are distributed all over the subcontinent without any linguistic preference (Kivisild et al., 2003a,b; Metspalu et al., 2004; Thangaraj et al., 2005a; Sun et al., 2006; Chandrasekar et al., 2009; Ref. I–IV).

Consistent with mtDNA results, the South Asian paternal lineages (Y-chromosome) are derived from the pan-Eurasian founder haplogroups CF and K (Sengupta et al., 2006; Underhill and Kivisild, 2007; Karafet et al., 2008). The haplogroup diversity among South Asian populations was found to be higher when compared to European or East Asian populations (Trivedi et al., 2008). The majority of Y gene pool of South Asia are largely restricted to Indian subcontinent and are composed of haplogroups C, F*, H, J, L, O and R (Sahoo et al., 2006; Sengupta et al., 2006). The high frequency of paternal haplogroups *viz.* C5, F*, H, L and R2 associated with high STR variance has usually been considered to their indigenous origins in the subcontinent (Kivisild et al., 2003a; Sahoo et al., 2006; Sengupta et al., 2006). However, the wide geographical spread of haplogroup R1a and the absence of sufficient downstream SNP marker make it difficult to infer its exact geographic origin. The spread of haplogroup O2a is characteristic of the Austroasiatic speaking populations of India and South East Asia (Kayser et al., 2006; Kumar et al., 2007).

The Y-chromosome as well as mtDNA studies clearly reflects the long term genetic and geographical isolation of Indian populations with minor influx from outside (Kivisild et al., 2003a; Metspalu et al., 2004; Thangaraj et al.,

2005a,b,2008,2009; Sahoo et al., 2006; Sengupta et al., 2006; Sun et al., 2006; Chandrasekar et al., 2009). Genetic distinction between the geographic sub-regions and origin of several deeply rooted haplogroups within the Indian subcontinent can only be dissected at the finer level of subclades of the mtDNA as well as Y-chromosomal haplogroup based on novel SNP discovery and complete sequence data at higher resolution. Therefore, there is a need now to delineate the phylogeographic pattern of such lineage within the subcontinent to infer the population's Neolithic settlement in the light of evidence from archaeological discoveries (Fuller, 2007). Such questions are now demanded to infer deep knowledge about the regional specific migrations in South Asia.

2.4.3.1. Rare polymorphisms

Sequence variation in mtDNA has been widely used to assess genetic relatedness at both species and population levels. In the consensus sequences of human mtDNA, there are two copies of the 9bp motifs "CCCCCTCTA" in the non-coding region-V. Deletion of one copy has been reported in many populations worldwide at different frequencies (Thangaraj et al., 2008 and references therein). In central and East Asian population 9bp deletion is predominantly associated with haplogroup B (Hertzberg et al., 1989; Yao et al., 2000). The 9bp indel polymorphism has also been extensively studied in South Asian populations and a lower frequency has been observed (Watkins et al., 1999; Clark et al., 2000; Thangaraj et al., 2005a,b; Kumar et al., 2006b), suggesting their independent origin in different haplogroup background, unlike to East and Southeast Asians. The first large scale study on Indian subcontinent provided evidence about different maternal origin of Munda and Mon-Khmer speakers (Thangaraj et al., 2005b). It was also supported by another similar study (Kumar et al., 2006b). These findings suggested that the Austroasiatic populations of the Indian subcontinent and Nicobar Islands have distinct genetic maternal sources. Another extended study (Thangaraj et al., 2008) refuted the theory of natural selection in shaping the mtDNA distribution in South Asia (Mishmar et al., 2003; Ruiz-Pesini et al., 2004). Moreover, this study also reported a new South Asian specific basal lineage M43 and suggested that geography of Nepal has received a bidirectional gene flow from East as well as West. A recent study on larger population size on Tharu population reported B5a lineage with 9bp deletion, bolstering above conclusion (Fornarino et al., 2009).

Another rare polymorphism present in South Asia is Y-alu polymorphism (YAP). The YAP polymorphism has been one of the most widely surveyed human polymorphisms. It originated with insertion of an *Alu* repetitive element at the DYS287 locus at location Yq11 (Hammer et al., 1997). Studies suggest that South-Asia contains both the major lineages (haplogrop D-Asian specific and haplogroup E-African specific) of this polymorphism and they entered at different timescales (Thangaraj et al., 1999,2003a; Qamar et al., 1999; Agrawal et al., 2005; Sengupta et al., 2006; Chandrasekar et al., 2007; Fornarino et al.,

2009; Eaaswarkhanth et al., 2010). Frequency of YAP+ E lineages is highest among sub-Saharan African populations (82–95%), followed by North African populations (50–70%), Middle East and Central Asian populations (12–27%) and less frequent (<10%) among Europeans and Asians (Hammer et al., 1998; Underhill et al., 2001; Quintana-Murci et al., 2001; Jobling and Tyler-Smith, 2003; Al-Zahery et al., 2003). Asian-specific YAP+ D lineages are frequent among Tibet, Japan, Himalayas and Andaman islands and are infrequent throughout East and South Asia (Thangaraj et al., 2003a; Hammer et al., 2006; Gayden et al., 2007). Based on high frequency of haplogroup D in Andaman Islanders, it was suggested that this haplogroup is the remnants of Out-of-Africa migration (Thangaraj et al., 2003a), but nearly absence of this lineage in Indian subcontinent (Sahoo et al., 2006; Sengupta et al., 2006), as well as in the light of new mtDNA evidence (Barik et al., 2008), it is likely that this haplogroup arose later in the East of Indian subcontinent after OOA migration. So far, three main sources (two for E and one for D haplogroup) have been identified for the presence of this polymorphism in South Asia. First E lineage-recent entry of Siddi population as slave trade (Thangaraj et al., 1999; Ramana et al., 2001; Gutala et al., 2006), second E lineage-through invasion of Muslims between 10^{th} and 16^{th} centuries from Central Asia and Iran (Qamar et al., 1999; Agrawal et al., 2005; Sengupta et al., 2006; Eaaswarkhanth et al., 2010), and third haplogroup D entrance towards Himalayan fringes from Tibet (Fornarino et al., 2009).

2.4.3.2. Pioneer settlement in Andaman Islands

The origin of the Andaman islanders has been a subject of speculation since a long time. The isolated populations of these islands were one of the most unique inhabitants of earth upholding the important clues to the evolution and dispersal of early modern humans. Because these enigmatic populations share physical features with African pygmies, including short stature and dark skin, and contrast with other Asian pygmoid people, there has been speculation that they may be directly related to African pygmies (Howells, 1974). The archaeological records about the Andaman Islanders is scarce and does not extend beyond the first millennium BC (Cooper, 2002). Later, two parallel studies, one dealing with mtDNA studies on ancient remains of these populations (Endicott et al., 2003), while another on modern DNA using mtDNA and Y-chromosomal markers suggested their affinity with Asian populations (Endicott et al., 2003a,b; Thangaraj et al., 2003a). Subsequent high resolution research using complete mtDNA data from the Onge and Greater Andamanese populations defined the two Andaman specific clads M31 and M32 suggesting them as the remnants of a single rapid dispersal pattern along the coast of the Indian Ocean during the late Pleistocene (Thangaraj et al., 2005a,b). In contrast to this, the finding of deeply rooted branch of M31 lineage in mainland India suggested origin of this lineage in mainland India (Palanichamy et al., 2006). The increased number of complete sequences from mainland as well as from

Andaman islands corrected the phylogeography and emphasized that M31 and M32 haplogroups are the sister groups within a single haplogroup, M31'32 rooted with the main trunk of M with a rare polymorphism (2156+A) (Barik et al., 2008). This study again revealed either of both assumptions *i.e.* the early (~60KYA) or (<25KYA) human settlement scenarios to this island. A much recent study dealing with 560,123 autosomal SNPs suggested their unique genetic identity and ancient isolation from mainland south Asian populations (Reich et al., 2009).

2.4.4. Austroasiatic dispersal: From east to west or from west to east?

The geographic origin and time of dispersal of Austroasiatic speakers, present in Southeast, East and South Asia, is still unclear. The Austroasiatic language family has been conventionally divided into three major branches- the Munda, the Mon-Khmer and Nicobarese (Diffloth, 2005; Blench, 2008). The recent classification catagorize Austroasiatics in to two major branches: Munda and Khasi-Aslian (Diffloth 2009). In South Asia three significant branches of Austroasiatic *i.e.* Munda, Khasian and Nicobarese, have been reported (see Fig. 5 of Ref-IV). Munda speakers are classified into Southern and Northern branches and dispersed in Eastern and Central part of India while Khasi reside in Meghalaya state of India and are the only representatives of Khasian branch. Nevertheless, most of the Austroasiatic speakers belong to the Khasi-Aslian group, dispersed in East and Southeast Asia. It was estimated that out of 90 million Austroasiatic speakers majority (70 million) of speakers belong to Vietnamese, 10 millions speak Khmer and 5 millions speak Santhali, while remaining 150 odd languages are spoken by several population groups, ranging from few to several hundreds or thousands of speakers (van Driem, 2001). Nicobarese live in Nicobar islands (a part of Southeast Asia, geographically) and speak six different types of Nicobarese language (Radhakrishnan, 1981). The migration of Nicobarese was thought to be associated with the agricultural expansion (Blench, 2008). It was also supported by the abundance of the savannas of *Imperata cylindrica* grasslands over main islands suggesting forest clearance by incoming agricultural populations (Blench, 2008). The genetic as well as linguistic studies suggest a closer affinity of Nicobarese with Island Southeast Asia (van Driem, 2001; Prasad et al., 2001; Thangaraj et al., 2003a ,2005a,b; Blench and Dendo, 2007; Blench 2008).

The agricultural expansion and animal domestication are two incidents which facilitated Anatomically Modern Humans (AMH) to expand rapidly. On the basis of single domestication event of rice (*Oryza sativa*), it was suggested that the ancestors of AA speakers scattered from the Yangtze River basin (South China) and moved out from primary agricultural homeland (Higham, 2003; Blench, 2005). The alternative model based on multiple domestication events of rice (Fuller, 2007) and comparative phonological evidence supports an Eastern

Indian (Orissa state) origin of this language group (Witzel, 2005). Hence, the domestication event and homeland of rice has been debated since a long time and is still in controversy (Kovach et al., 2007; Fuller, 2007; Vaughan et al., 2008). The cultivation of rice is distributed in five continents. It is one of the major food crop and carbohydrate suppliers for Asians, Africans and South Americans although, its domestication event remains still obscure (Khush, 1997; Kovach et al., 2007; Fuller, 2007). Two main rice species known are Asian rice (*Oryza sativa*), which is cultivated worldwide and African variety (*Oryza glaberrima*), exclusively present in Africa. It has been clearly established that both of these species originated from two discrete domestication events. The domestication event information about African rice is straightforward while the major debate now is about Asian rice. It is well-known that the closest wild relative of the Asian species *Oryza sativa* is *Oryza rufipogon*, a species found solely throughout Asia, however, the main debate concerns the number of domestication centers. Present day, there are two main types of *Oryza sativa i.e. japonica* and the *indica* types.

Several scholars have proposed that these two forms may have originated from two distinct domestication events (*e.g.* (Kovach et al., 2007; Fuller, 2007; Vaughan et al., 2008). Though there are few archaeological records of rice cultivation in East Asia, including Korea and Taiwan explored recently, but records from South Asia related to rice domestication is poor due to tropical climatic conditions. The archaeological records from East Asia argue that there is only one center of domestication of rice in Asia located in the Yellow river basin which later differentiated through selection into the two main cultivar types known today, *i.e. japonica* and *indica* (Goff et al., 2002; Higham, 2003; Blench, 2005). Since traditional *japonica* varieties are predominantly found in East Asia, and *indica* varieties are predominantly found in South Asia, the alternative hypothesis conjectures that the j*aponica* and *indica* types originate from two distinct domestication events (Fuller, 2007).

The completion of the rice genomic sequence, in particular, has been considered as a milestone in agricultural research because it gave access for the first time to the complete gene repertoire of a crop species (International Rice Genome Sequencing Project, 2005). Molecular studies on rice varieties illustrated the genetic differentiation of both *indica* and *japonica* type (Glaszmann, 1987; Wang et al., 1994; Prashanth et al., 2002; Garris et al., 2005). It should be noted that this genetic differentiation into two distinct gene pools is not in conflict with the single domestication hypothesis, because it could be the result of a strong artificial selection for the two distinct plant types, posterior to the domestication. Recently, two studies based on analysis of transposable elements and molecular clock estimation suggested that that *indica* and *Japonica* types arose from two distinct domestication events (Vitte et al., 2004; Ma and Bennetzen, 2004). The MRCA of *Japonica* and *indica* was calculated much prior to their domestication event (Vitte et al., 2004; Ma and Bennetzen, 2004). Genome-wide studies of genetic variation demonstrate that the two varietal groups *i.e. indica* and *japonica* in *Oryza sativa* arose from genetically distinct

gene pools within a common wild ancestor, *Oryza rufipogon*, suggesting multiple domestications of *O. sativa* (Kovach et al., 2007 and references therein). More recently it was reported that PROG1 (PROSTRATE GROWTH 1) gene controls wild-rice plant architecture (Jin et al., 2008). The gene mapped on chromosome 7, encodes a newly identified zinc-finger nuclear transcription factor with transcriptional activity. Transformation experiments demonstrate that artificial selection of an amino acid substitution in the PROG1 protein during domestication disrupts the PROG1 function and inactivates PROG1 expression, leading to erect growth, greater grain number and higher grain yield in cultivated rice (Jin et al., 2008). Sequence comparison shows that *indica* and *japonica* varieties of rice carry identical mutations in the PROG1 coding region suggesting a single rice domestication event (Jin et al., 2008; Tan et al., 2008). Therefore, recent genetic studies in contrast with previous studies support a single East Asian domestication of rice although this needs further exploration in terms of it's relation with AA speakers. More importantly, it should be noted that many AA populations (Munda speakers) of South Asia are hunter-gatherers, and therefore languages are not typically diffused by sedentary agriculturalist populations but by mobile groups, where the immigrants form a numerically strong underclass (van Driem, 2001). Hence, it is problematic to associate the rice farming with AA expansion.

Two contending hypotheses have been proposed, one of which places the origin of Austroasiatic speakers in Southeast Asia with a later comparatively recent dispersal to South Asia (Diamond and Bellwood, 2003; Sahoo et al., 2006), whereas the second hypothesis advocates Indian origin of the populations of this language family (Roychoudhury et al., 2001; Basu et al., 2003; Kumar et al., 2007; Fuller, 2007). Previous genetic studies (Kivisild et al., 2003a; Basu et al., 2003; Kayser et al., 2006; Kumar et al., 2007) have revealed region-specific patterns of mtDNA contrasted with across-regional spread of one particular Y-chromosome haplogroup, O2a, providing the first genetic correlate for the spread of this language group (Fig.8). Haplogroup O2a (M95) is most frequent among Munda speakers (Basu et al., 2003; Kumar et al., 2007). The existence of relatively high frequency of M95 (O2a) in both Indian (Munda) and SE Asian populations is intriguing given that previous studies show substantial maternal genetic isolation between these two regions (Fig. 8a) (Metspalu et al., 2004; Black et al., 2006; Ref II–IV). The mtDNA information available so far insinuates clear distinction of Indian Munda and SE Asian Mon-Khmer speaking groups, with former sharing of their basic mtDNA haplogroup structure with other Indian populations (Fig. 8a) (Basu et al., 2003; Metspalu et al., 2004; Black et al., 2006; Kumar et al., 2006a,b; Reddy et al., 2007; Ref. I–IV). Consistent with their linguistic separation, the Mon-Khmer speaking Nicobarese carry exclusively East Asian specific mitochondria (Prasad et al., 2001; Thangaraj et al., 2003a). Notably, Khasi (Mon-Khmer) group, residing in Meghalaya state in India, shows admixed package of both Indian and East Asian lineages (Reddy et al., 2007).

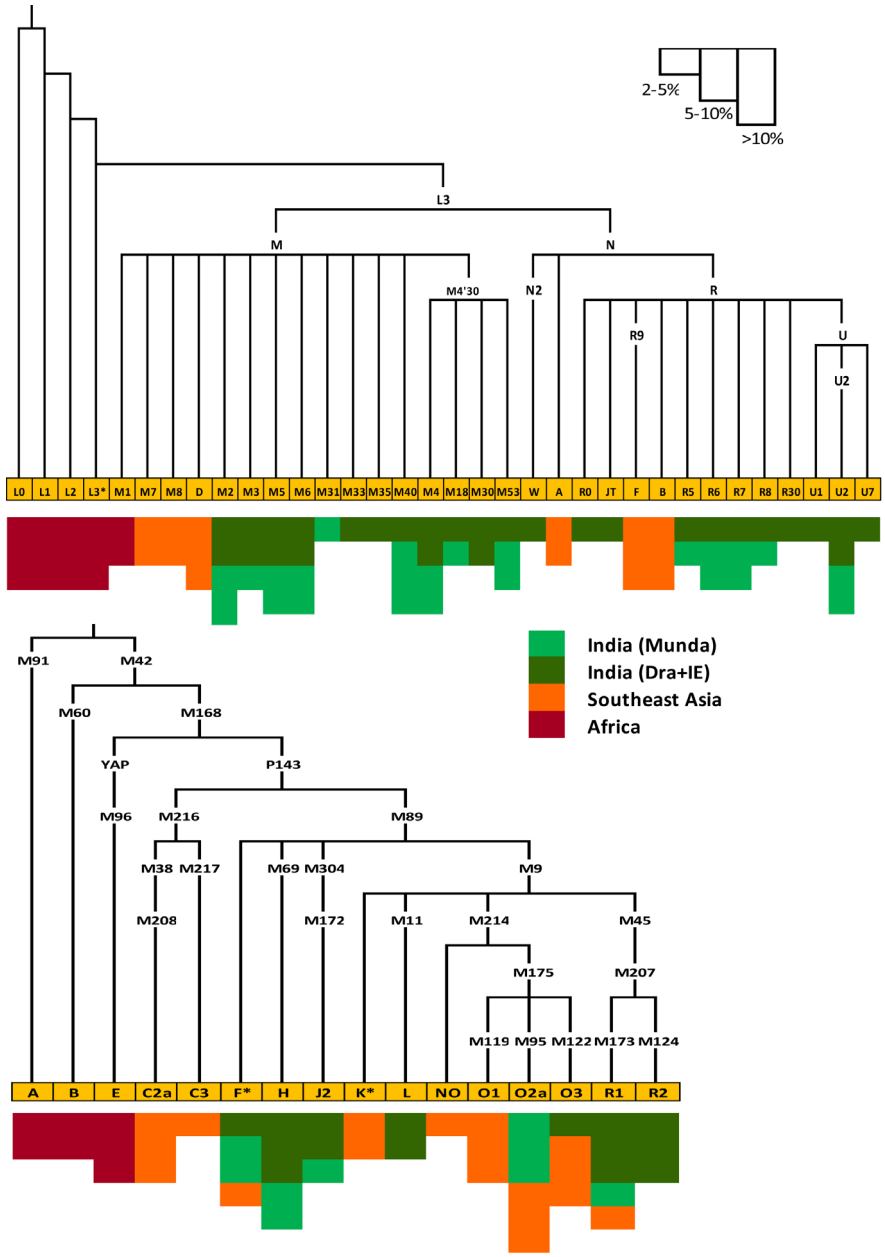

Figure 8. The structuring of different mtDNA (upper pan) and Y-chromosdomal major Haplogroups (lower pan) in SE Asians and Indian populations. The Y-chromosomal monophyletic haplogroup O2a is common and prevalent in both SE Asian as well as Indian (Munda) populations (b), while, in case of mtDNA regional specific branches emerge directly from the pan Eurasian founder haplogroups and have no such nested composition (a). The mtDNA information available so far insinuates clear distinction of Indian Munda and SE Asian Mon-Khmer speaking groups, with former sharing of their basic mtDNA haplogroup structure with other Indian populations.

In contrast, lack of common mtDNA lineages among Mon-Khmer and Munda, the Y-chromosome pools of Indian and Southeast Asian AA speaking populations share a common marker, M95, which defines a single branch (O2a) of the otherwise strictly East Asian specific tree of haplogroup O (Fig. 8b). The distribution of the hg O2a (M95) underscores that this Y-chromosomal SNP is exceptionally informative with regard to the genetic link between AA populations of SE and South Asia. Using STR data for 34, O2a chromosome samples from two North-Munda speaking groups (Ho and Santhal) of India and applying evolutionary mutation rate $6.9x10^{-4}$ mutations per generation per locus (Zhivotovsky et al., 2004; Sengupta et al., 2006), the coalescent age of M95 lineages among Indian AA speakers was estimated as 8.8 ± 2.2 KYA. Similar 8.8 KYA coalescent date, using substantially higher germ-line mutation rate of $2.8x10^{-3}$ mutations per generation per locus, was obtained for SE Asian M95 lineages earlier by Kayser et al. (2003). Bearing in mind the differences in mutation rates, these estimates indicate that M95 diversity in Indian Munda speakers is considerably lower than that observed in Southeast Asia. Surprisingly, however, another survey of Indian AA populations, covering in total 25 populations, reported significantly higher coalescent (65 KYA) time of Indian M95 using the evolutionary mutation rate and an extended list of 20 STR markers, suggesting deeply rooted Indian origins of haplogroup O2a (Kumar et al., 2007). Nevertheless, no direct comparison with East Asian M95 chromosome was provided via the same STR markers employed in this study. It should be noted that the 65 KYA age estimate of haplogroup O2a in India appears much older than the estimated of the age of it's parental haplogroups K and NO (Rootsi et al., 2007; Karafet et al., 2008).

If indeed O2a originated in India and later migrated to SE Asia, then we should also observe sister clades of hg O2a, such as O1 (M119) and O3 (M122) in India, which is not the case except for Sino-Tibeten populations. O1 (M119) and O3 (M122) are limited to Northeastern part and are descendants of recent migration from East (Cordaux et al., 2004b; Metspalu et al., 2004). Instead, we do observe a splendid scattering of these sub-clusters in Southeast and East Asia (Kayser et al., 2006; Xue et al., 2006; Li et al., 2007; Li et al., 2008). Furthermore, NO and N clades are also not reported in India, whereas in East and Southeast Asia they are present in notable frequency (Sahoo et al., 2006; Sengupta et al., 2006; Xue et al., 2006; Rootsi et al., 2007; Ref. I–IV). However, we cannot pinpoint exactly the precise geographic source of hg O2a as we don't have any information about hg O2a (M95) in AA speaking populations of Mynmar, which is a likely source for the ancestral populations of Munda speakers of India. However, the available evidence indeed suggests the Mainland SE Asia as an early and important geographic location for hg O2a. More data from East, Southeast and South Asia of haplogroup O2a is required to resolve the origin of this mysterious haplogroup as well as AA homeland. Therefore, present Munda speakers are unlikely to be the source population for hg O2a (M95) and thus the model supporting Indian origin of hg O2a doesn't fit with the current topology.

2.4.5. The newcomers: Whole genome approach

Mitochondrial and Y-chromosomal datasets can provide insights into population structure, but each effectively assays only a single locus resolution in a sex-specific pattern and is more prone to drift and other population selective pressures. Therefore, population structure and its genetic diversity play an important role to understand the nature and extent of any disease specific association studies that aim to discover genetic factors implicated in human health and disease. The adaptation of humans in a particular environment often leads to an increase in beneficial gene which is known as positive selection. The positive selection increases the frequency of favorable genes in a particular population resulting in high level of population differentiation or skewed allele frequency and leaves an imprint on the pattern of genetic variation found in a population near the site of selection. Such pattern can be recognized by comparing the DNA variants in multiple individuals from the same population and comparing them with other adjoining populations (Voight et al., 2006; Sabeti et al., 2007). It is evident from genetic studies that positive natural selection has acted on several regions of genome (Bryk et al., 2008; Sabeti et al., 2007; Soejima et al., 2006; Izagirre et al., 2006; Bersaglieri et al., 2004; Sabeti et al., 2002; Xue et al., 2009b; Voight et al., 2006; see also Hurst, 2009). However, a number of studies have also emphasized the fact that large differences in allele frequency between populations may not always be necessarily due to positive selection (Gardner et al., 2007; Hofer et al., 2009).

The International HapMap populations studies have provided valuable information on inter-continental variation across the human genome, including structural variation, recombination and selection on Yoruban, Japanese, Chinese and European-Americans populations (International HapMap Consortium, 2005,2007). This project provided fine scale variation at the resolution of more than 3.1 million SNPs genotyped in 270 individuals from three major continental populations and also highlighted 32 SNPs from 27 genes that exhibited particular evolutionary interest because of their non-synonymous nature and high level of population differentiation. Moreover, knowing the fundamental genetic structure of any population being scrutinized for association studies is the most important factor to consider, as the ignorance of underlying genetic structure of a population can lead to a false signal. Hence, a clear understanding of the genetic structure of human populations is fundamental to medical science.

Thus, multiple autosomal loci provide a very robust and high resolution assessment of population structure as well as disease association (Jakobsson et al., 2008; Salmela et al., 2008; Li et al., 2008; Heath et al., 2008; McEvoy et al., 2009; Nelis et al., 2009; Xing et al., 2009; Auton et al., 2009; Reich et al., 2009; HUGO Pan-Asian SNP Consortium 2009; Behar et al. 2010). Another advantage of using large number of autosomal loci is in forensic purposes, where the geographic origin of unknown sample, even from an admixed population, can be assigned with the highest level of accuracy (Heath et al., 2008; Lao et al., 2008; Novembre et al., 2008). These studies along with others

demonstrated the possibility to make clear differences amongst individuals from closely related populations and even to observe stratification within the same population (Heath et al., 2008; Lao et al., 2008; Novembre et al., 2008; McEvoy et al., 2009). The first clustering approach was developed by Rosenberg et al. (2002) using >350 microsatellite sequences and identified five continental populations structure. These clusters arise from genuine features of the underlying pattern of human genetic variation, however, see Weiss and Long (2009) for the criticism. Such structure emerges with accumulation of small but concordant differences in allele frequency across many loci.

The whole genome approach signifies the existence of substantial genetic structure among samples from different continents (Li et al., 2008; Heath et al., 2008; McEvoy et al., 2009; Nelis et al., 2009; Xing et al., 2009; Auton et al., 2009; Reich et al., 2009; HUGO Pan-Asian SNP Consortium, 2009; Behar et al. 2010). In many populations, individuals predominantly cluster in only one group whereas in other populations, there are clearly multiple clusters. This mixed ancestry can arise from recent admixture among many founder populations or it can be due to shared ancestry before the divergence of subpopulations which is hard to interpret. But along with the data of haploid markers one can infer precisely if it is a result of recent gene flow or common ancestry before the divergence. For example, the East Asian ancestry component of Hazara (Li et al., 2008), comes from their Mongolian chromosome (Zerjal et al. 2003), the presence of sub-Saharan African component in Bedouins and Palestinians can unequivocally seen from their mtDNA pool (Li et al., 2008; Abu-Amero et al., 2008).

In spite of the fact that India comprises one fifth of the world population, neither the HapMap, nor the global panels that have been used for whole genome analysis included any population from India, though the inclusion of Pakistani populations did provide some clue about population structuring of South Asia (Jakobsson et al., 2008; Li et al., 2008). Previous population genetic analyses with autosomal markers in South Asia had a wide geographic but extremely low resolution of genomic coverage (Basu et al., 2003; Watkins et al., 2005; Reddy et al., 2005; Kashyap et al., 2006). Another study based on autosomal microsatellite markers had inferred that Indian populations show low levels of genetic differentiation (Rosenberg et al., 2006), which was however, criticized due to limited sampling and low geographical coverage (Indian Genome Variation Consortium, 2008). Later, the Indian Genome Variation Consortium (Indian Genome Variation Consortium, 2008) studied 1871 individuals from 55 diverse populations and provided a high degree of genetic differentiation among Indian ethnic groups. The sampling encompasses six geographical regions. Three ethnic categories were taken in to consideration: caste groups, tribal isolated populations and religious groups. Using this genetic data, they identified five clusters among these 55 populations. It should be noted that none of the cluster is 100% comprised with just a single component, but based on the larger proportions of the genetic component: first and second clusters comprise north Indian Indo-European and Sino-Tibetan populations group

respectively, third and fourth clusters consist Western Indo-European and Dravidian populations respectively, while fifth cluster harbors Munda populations (See Fig. 3 from Indian Genome Variation Consortium, 2008). This study also claimed a correlation of language with gene. On the contrary, studies based on haploid genome markers support the significant correlation between genes and geography (Kivisild et al., 2003a; Thangaraj et al., 2009; Ref. IV). The most parsimonious explanation of this discrepancy is inclusion of comparatively higher number of Austroasiatic and Sino-Tibetan samples (who have an overwhelming exceptional Y genepool than local Indian populations – O2 and O3 haplogroups, respectively) , than their population sizes (for instance, Austroasiatic and Sino-Tibetan represent 20% of the pooled samples comparing to their population size in India *i.e.* 1.8%), which has skewed the corralogram towards language and increased the correlation coefficient of gene with language rather than with geography. Furtheremore, it should be also noted that the four major linguistic families in India, by and large, have their own non-overlapping geographic domains which can counteract the correlation between gene and geography, thus providing an artificial correlation of gene with language.

Recently, couple of studies came up with few Indian samples showing caste-tribal differences and North-South gradient (Auton et al., 2009; Xing et al., 2009). The South Indian castes were suggested to have different ancestry proportions, Brahmins of South India have larger proportion of European component than tribals and dalits while later (*i.e.* tribal and dalits) share largely similar type of structure (Xing et al., 2009). More recently, a high profile genetic study analyzed 560,123 SNPs among 132 individuals derived from a diverse 25 ethnic Indian populations using the similar approach and identified two major ancestral populations *i.e.* ASI (Ancestral South Indian) and ANI (Ancestral North Indian) (Reich et al., 2009). The Onge branch seems to descend from an ancestral population which also gave rise to ASI while the ancestry of both of the components in other Indian populations varies between 39-71%. It is interesting to note that in this study despite many shared genetic features for caste and tribal South Asian populations, some populations have evolved group-specific patterns (forming outliers in PCA) of genetic diversity *e.g.* Onges and Chenchus. A similar, group-specific pattern is also reflected in Maya and Pima populations of South America (Li et al., 2008). Such structuring corroborates that long term isolation and genetic drift may have acted rapidly to produce detectable population-specific structure in small isolated populations. The analysis of population structure based on individual genotype suggests that some populations specifically have a pattern of reduced diversity, which is a characteristic feature of genetic isolates. These results are consistent with high levels of genetic drift and isolation among South Asian populations which indicate their importance as a potential candidate for association studies. The existence of sub-structuring in populations from Indian subcontinent has notable implications for population genetic studies and forensic databases where broad grouping of populations based on such affiliations is frequently employed.

3. AIMS OF THE PRESENT STUDY

South Asia constitutes a large segmented endogamous society that harbors rich genetic diversity within itself and offers to test several population genetic models derived by linguistic, geographical and cultural boundaries that can have a profound effect on human evolution. Strict endogamous populations, flourished independently with varied socio-cultural and linguistic diversities, nurtured by the vast geographical and ecological system, enriched South Asian diversity tremendously. It is of particular interest to study patterns of genetic affinities among endogamous groups inhabiting small geographical regions within the subcontinent because of their diverse origins and interethnic separation. The strict practice of endogamy across all social ranks resulted in emergence of population-specific diverse social traditions and development of distinct linguistic dialects. These divergent socially structured population groups provide a varied substratum for understanding variation of a genetic trait, spread of a particular disease and their prehistorical settlement in a geographical area. An understanding of evolutionary history of peopling of South Asia has long been a subject of interest for the researchers working in evolutionary and medical genetics.

During the commencement of this work the largely maternal autochthonous genepool of South Asia started gaining support due to higher resolution studies (Metspalu et al., 2004; Palanichamy et al., 2004) although, the skeleton of macrohaplogroup M, which encompasses 60% of Indian maternal genepool was not resolved completely. Besides this, the gene-culture interaction, caste-tribe continuum and localized haplogroup expansion were some of the basic questions. We therefore added high resolution molecular insights into genetic, linguistic and anthropological background and updated the phylogenetic knowledge about peopling of South Asia.

The questions undertaken for the present work are as follows:

- MtDNA haplogroup M comprises more than 60% of the matrilineal diversity of populations living in South Asia. Furthermore, a vast majority of them are autochthonous, not found or almost not found elsewhere in Eurasia, Melanesia and Oceania. Because such Indian M lineages differ profoundly from numerous M lineages found elsewhere, they offer unique insight into the matrilineal heritage of the subcontinent. Several previous studies have suggested that Indian mtDNA lineages, particularly those classified under haplogroup M, have evolved locally in India. However, substantial proportions of the M lineages in India have been studied so far only at low molecular resolution and involve insufficient sampling. Due to the scarcity of the data, it is not known, whether the so-called M* lineages detected at variable frequencies across India are more closely related to the local M subclades or to those widely spread in East Asia or Oceania. Is it possible to detect at complete mtDNA sequence resolution level elements of nested

phylogenetic structure that would link South and East Asian populations or either of the two more specifically with Papuans and Australians?

- Andaman islanders are classically considered as one of the most remote populations in the world of unknown genetic origin. Often called in anthropological literature as Negritos for their dark skin colour and curly hair, different hypotheses on their ancestry have been proposed. It was recently shown that two mtDNA haplogroups, M31 and M32, which occur at frequency <0.1% in mainland India, characterize almost all mtDNA lineages in Southern Andaman islanders. We further sought to examine whether the fine internal structuring in M31 and M32 haplogroups phylogeny reconcile with cultural and geographical separations of Andaman islanders?
- Since the period of classical genetic markers, Indian social stratification has been one of the most complicated issue to address. Many papers have been devoted to demonstrate differences between the castes, but much less attention has been addressed to numerous tribal people still living in India. And that irrespective of the fact that there have been hypotheses, based largely on indirect evidence, that the tribal people might be "true autochthonous populations" inhabiting India, while the speakers of not only Indo-Aryan but also Dravidian languages, are relative newcomers from the Neolithic period that covers only the last period since the subcontinent has been inhabited by modern humans. What is the reason behind the contrasting distribution of the scheduled castes and scheduled tribes in India? Is it due to social mobility governed by the caste-tribe continuum? Furtheremore, several previous non-genetic studies documented language shift as a common phenomenon among South Asian populations where several populations have very likely changed their mother tongue. However, none of the molecular studies have been so far aimed at testing this assumption. Therefore we have now asked whether the phenomenon of language shift typical for South Asia as such, or is it primarily a signal of a rapid genetic admixture?
- One particular, but otherwise representative example of ambiguity in the languages versus genes problem has been the origin of the Austroasiatic speaking people in India. Some influential authors still consider them as autochthonous, pre-Aryan and even pre-Dravidian inhabitants of India, others look to them as very recent newcomers from Southeast Asia. Therefore, it can be considered as a sort of test question for a geneticist – can one now, with all the new powerful tools to investigate phylogeography of mtDNA (and of the Y chromosome), find ways to answer these questions? Our preliminary analyses of mtDNA haplogroup R7 at low resolution (HVS-I) showed it's significantly higher frequency in Austroasiatic (Munda) speaking tribal populations than among other language groups of India. Therefore, a higher molecular resolution study was designed to test if R7 can be considered as an Austroasiatic (Munda) specific haplogroup?
- Like elsewhere in a global context, genetic boundaries and differences in general, among populations of South Asia, are likely governed by several

factors, such as geography, language and culture. Because of its geographic span and, in particular, extremely complex social stratification and also the size of population, we found it important to seek an answer (even if only as a preliminary attempt) to the most general question: Is the South Asian genetic landscape primarily molded by interaction of genes and languages or genes and geography?

4. RESULTS AND DISCUSSION

To find a reasonable explanation to each of the above questions we tested the genetic diversity, extent of geneflow across cultural boundaries among several ethnic groups of South Asia and evaluated the interaction of gene-language-geography on structuring of populations. We have also reconstructed some of the suggested population histories from a molecular perspective. Altogether, more than 12,000 samples were analyzed for specific genetic markers. The informed consent was obtained from all participants.

4.1. Indigenous origin of deep-rooting South Asian maternal lineages (Ref. I)

In this study 8670 DNA samples from more than 150 ethnic groups from India were genotyped for mtDNA haplogroup defining markers and sequenced for the first hypervariable region (HVS-1) of the D-loop. Majority of the samples could be assigned to previously known Indian specific sub-clades, with macro-haplogroup M accounting for more than 60% of the observed variations. However, a small proportion (18%) of the samples could not be designated specifically to any sub clade of haplogroup M and were thus referred as M*, as a convention to refer such lineages. Because haplogroup M sub-clades are found across a wide geographic area extending from East Africa to East Asia and Australia, we decided to test whether the M* lineages that we detected at low molecular resolution are related to any existing M sub-haplogroups in India or elsewhere. We determined the complete mtDNA sequence in case of 11 and partial coding sequence in case of 2231 of such Indian M lineages that either belonged to M* or whose phylogenetic affiliation to existing mtDNA haplogroups could not be accurately determined. The targeted samples also included Andamanese specific lineages from haplogroup M31.

The resulting high resolution data allocated us to define one novel sub-clade of M that we proposed to label as haplogroup M41. Based on the newly obtained sequence data we revised the classification of some major haplogroups, such as M3, M18, and M31 that stem directly from the root of haplogroup M. Our analyses enabled us to confirm and revise the substructure of a number of previously ascertained haplogroups, such as M3a, M4a, M6b, M33a, M34a, M37a and M40a. Our results confirm the local origin and diversification of the majority of Indian mtDNA lineages belonging to haplogroup M using significantly increased sample sizes (8670) and molecular resolution. We observed no significant differences among different language groups in the distribution of the major sub-clades of M lineages, although certain geographical patterns could be detected.

Above results suggested the autochthonous origin of these deeply rooted haplogroups which are not language specific and are spread over all the

language groups in India. Our reanalysis of the Andamanese-specific lineage M31 revealed two clear-cut population specific branches-M31a1 present in Onge and Jarwa individuals while branch M31a2 was exclusively found among Great Andamanese individuals, consistent with their linguistic separation. The starlike sprouting of South Asian specific lineages coupled with non-nested structure of the phylogeny (Fig. 2. Ref. I) support the one wave, rapid dispersal theory of modern humans along the rim of Asian coast.

4.2. Major factors which shaped South Asian genepool and caste-tribe continuum (Ref. II)

Throughout its history, South Asia has been a blending pot of ethnic groups, languages and cultures**.** There is a fundamental resemblance in various rituals practiced by people in different regions and therefore, shared ritual patterns can account for some unity in the varieties of the religious beliefs that we observe in South Asia for a long time. In ancient South Asia, there were several kingdoms living out of their own story of conquest and collapse although, these local dynasties were built upon the roots of a culture which was well established. South Asia has always been simply too large, too complex and too culturally subtle to let any empire dominate it for long. Therefore, it is not surprising to find Indian population genetically and socio-culturally as highly diverse as of today.

In this study we have extensively discussed various highly debated issues alongwith several evolutionary and social forces that have shaped the present form of South Asia *e.g.* caste system, interrelations among caste and tribes, Aryan invasion theory, origin of different language groups, earliest settlers and genetic architecture. The caste system of South Asia is maintained by several traditional rituals and social practices, including strict endogamy. Besides the scheduled castes (SC) population, the scheduled tribal population (ST) represent the historically most backward and disadvantageous groups in the highly ripped Indian caste society. While the social, economical and educational deprivation of these groups has been in the past a common and unifying characteristic, each group had and still largely has its own particular rituals that distinguishes one from the other. One important aspect of Indian caste system, which makes a substantial impact on the inferences one can make from the caste/tribal genetic variation, is that the definition of the "caste" has been historically fluid. Due to the influence of ruling authorities the definition of "caste" or "tribe" has constantly been changing since 500 years or more. Precisely, the scheduled tribes were gradually incorporated into the caste system as scheduled castes. Deforestation catalyzed the assimilation of hunter-gatherer societies into agriculturally based subsistence economies, thus, changing their label from scheduled tribe to scheduled caste. This dynamic process of absorption of the tribal population into the caste society explains the contrasting geographical distribution of Scheduled caste and Scheduled tribes in India (Fig.2. Ref. II).

In debate on single *vs.* multiple dispersals out of Africa, recent mtDNA and Y-chromosomal studies along with archaeological discoveries support a single early migration that brought ancestral mitochondrial (M, N and deriving from the latter, R) and Y-chromosomal (C, D and F) lineages to Eurasia and Oceania, suggesting single wave of their co-migration along the Southern route (Thangaraj et al. 2005a; Macauley et al. 2005; Hudjashov et al. 2007, Ref. I). The internal structure of haplogroup M and N (N1d and N5, R5–R8, R30, R31, U2a,b,c) lineages in India, however has revealed now by the analysis of complete mtDNA genomes, their basically autochthonous development (Ref. I,II,IV). The high STR variance and widespread existence of Y-chromosomal haplogroups C5, F*, H, R2 and L in Indian subcontinent have been considered as indicative to their indigenous origins. It is clear from genetic studies that certain genetic variants were found to be shared betweenIndian and European populations which shared a common genetic ancestry in late Pleistocene, suggesting gene flow into India during the period of the proposed Aryan invasion to be minimal. Few studies have suggested haplogroup R1a, as a potential marker for Indo-Aryan invasion for introducing the caste system to India as the frequency of this haplogroup was found specifically higher among the caste groups. The higher variance of STRs in Indian R1a lineages as compared to those from Central Asia further weakens such scenario implying a strong founder effect. However, the current lack of sufficient SNP marker resolution makes it difficult to infer the geographical origin of haplogroup R1a.

We have also taken into account several studies suggesting Austroasiatic groups as the earliest settlers of Indian subcontinent. However, as argued in Metspalu et al. (2004), it would be highly problematic to relate any haplogroup with the language family, because the language families involved are generally believed to be by far younger than the time frame relevant for peopling of South Asia. Moreover, such autochthonous deeply rooted South Asian lineage groups are widely spread across language borders in the subcontinent (Ref. I), and, as exemplified by the Mushar group (Ref. III), documented and putative language shifts make it difficult to infer the original mother tongue for every population studied even during the historic period and perhaps impossible for ancient times. Thus, the present-day linguistic affinities of different Indian populations *per se* are perhaps among the most ambiguous and even potentially controversial lines of evidence in the reconstruction of prehistory of South Asia.

4.3. Testing the model of "language shift" in South Asia (Ref. III)

Two contrasting scenarios have been proposed for the spread of culture and language in South Asia – the demic diffusion model – prevailing mass movement of populations with genetic exchange, and the cultural diffusion model - implying the dispersal of cultural-traits (*i.e.* language, agriculture *etc.*) without considerable migration of people between populations, causing limited

or no genetic exchange among them (Cordaux et al., 2004a; Sahoo et al., 2006). With the exception of some marginal groups, demic diffusion scenario however, was not supported in further high resolution studies with larger sample size (Sengupta et al., 2006; Ref. III). Language shift is documented in history - *e.g.* in British Islands, in Roman-time Sardinia, in Armenia after the fall of Urartu, in Anatolia with the arrival of Turks (Cavalli-Sforza et al., 1994) all of which have been examples of only limited genetic change. Few other prehistoric examples are the adoption of Austronesian languages by some former speakers of Papuan languages in the Western islands of Melanesia and by Aeta in the Philippines (Cavalli-Sforza et al., 1994; Kulick, 1998). It is a process where a dominant language is adopted by a contemporary population, with only a minor contribution of genes from a source population. In such cases persons belonging to minority population, in order to get a higher status in the society gradually shift their language from their mother tongue towards the language of majority host population. Language shift may be also forced by ruling authorities. However, the original language doesn't disappear quickly, it takes rather several generations. A prehistoric language shift from Uralic to Indo-European among the Baltic-speakers has been also hypothesized (Wiik, 2002).

The present language distribution of South Asia provides an opportunity to test various models of language dispersal, language shift being one of them although there is a lack of detailed studies of language shifts in South Asia. Therefore, we tried to investigate the correlation between language and genes in a population called "Mushar" who have changed their language from Austroasiatic (North Munda) to Indo-European (Bhojpuri, Maithili, Magahi, Bengali Bundelkhandi, Nepali, Chattishgarhi *etc*. dialects) in the recent past. The word "Mushar" has been derieved from the word 'Mush', a Bhojpuri dialect word meaning mouse or rat and 'ahar' (Sanskrit word) meaning eaters thus referring them as "Mouse eaters". They live mostly in Eastern India extending themselves to Central and Northern parts till Nepal (Fig. 1, Ref. III). According to the elderly members of the Mushar community they have migrated from the East to different parts of the country and further to Nepal. It is interesting to note that in some of the Indian states they are considered as a scheduled caste (SC) and in some others as a scheduled tribe (ST) and in pre-independence period, they were unanimously classified as ST (Dungdung, 2003), which again reflects the phenomenon of caste-tribe continuum (Ref. II). They were hunters and gatherers initially but now live in villages and have adopted agriculture. Most of them currently speak Indo- European language but few (older people) are still bilingual, speaking Indo-European and Austroasiatic (North-Munda) as well. To address this question, we analyzed maternal (mtDNA) and paternal (Ychromosome) lineages of the target population and compared the results with those of the juxtaposed Austroasiatic and Indo-European populations.

The maternal as well as paternal genetic analysis of Mushar population along with their neighboring populations revealed a haplotype sharing with both Austroasiatic and Indo-European speakers. However, most of the haplotypes

sharing was with Austroasiatic populations. We have also identified two novel indigenous mtDNA haplogroups M45 and M46 (Fig. 2, Ref. III). Besides this, we observed that the Austroasiatic (Munda) specific haplogroups such as M40a, M45, R7 and R6a are also more frequent in the Mushar population and these are exclusively shared with the surrounding Austroasiatic populations considered for this study (Fig. 2, Ref. III). Most of the Indo-European samples don't share haplotypes with either their surrounding Austroasiatic or Mushar populations. The Y SNP tree also illustrates that as in the case of neighboring Austroasiatic populations, O2a (M95) is the most frequent haplogroup in the Mushar population (Fig. 3, Ref. III). The Analysis of Molecular Variance (AMOVA) based on mtDNA and Y-SNPs (Table 2, Ref. III) suggests that the Mushar is quite almost four-fold differentiated from the Indo-European populations.

We have also tested the assumption that, "a caste having the same name may have different genetic makeup" at different geographical regions (Karve and Malhotra, 1968; Majumder, 2001b) by exploring the genetic structuring of Mushar populations collected from different geographical regions (Fig. 4, Ref. III). The largely uniform composition of maternal as well as paternal genepool suggests a common paternal and maternal origin of these groups even though they inhabit distant geographic regions. Therefore, "a caste having the same name may have different genetic makeup" assumption can't universally applied on the South Asian populations and it also depends upon social uplifiting process (also referred as *Sanskritization*) (Ref. II, III). The most plausible cause of language shift in Mushar was their assimilation into the agricultural community where the agricultural land largely has been the property of Indo-European speaking communities. To get the work and food in such conditions learning Indo-European language was the most advantageous source of their survival. Moreover, the education of their children in the Indo-European schools added the impact to rapid loss of their original language.

4.4. Phylogeography of mtDNA Haplogroup R7 in India: the demise of a simple diagnosis (Ref. IV)

South Asia is an ideal region for studying the relationships between culture, geography and genes and for developing interdisciplinary approach concerning the demographic history of anatomically modern humans (AMH). Besides this, South Asia was probably the major outcome of the dispersal of AMH Out-of-Africa as attested by large number of deep-rooting mtDNA lineages sporting from the basal nodes of both superhaplogroups M and N (including R). These deep-rooting mtDNA haplogroups are spread over the cultural and social borders testifying to the general autochthonous diversification of the maternal gene pool of Indian subcontinent and hence, the common origin for the people that speak diverse languages today and belong to different castes and tribes

(Ref.I, II, III). Consequently, the haplogroup richness of the Indian subcontinent appears to have formed *in-situ* and to date back to Late Pleistocene, approximately 40,000–60,000 years back. The early demographic expansion just after the initial colonization and succeeding *in-situ* origin of several deep rooting lineages has founded the diversity of modern humans in South Asia. Comparisons of relative regional population sizes through time, deduced from Bayesian coalescent inference methods applied on mtDNA complete sequence data from global populations have suggested that approximately between 45 and 20 KYA most of humanity lived in Southern Asia (Atkinson et al., 2008).

The first extensive study of complete mtDNA sequences from India (Palanichamy et al. 2004) identified and solidified numerous indigenous clades (and sub-clades) emerging directly from the roots of superhaplogroups N, R and U, such as N5, R5-R8, R30, R31, U2a-d and U7. West and East Eurasian specific haplogroup families HV, JT, N1, and U (xU2a-d, U7), present at lower frequencies, suggest limited gene flow into Indian subcontinent, likely from West and Northwest Eurasia. In the present study, we extended the complete mtDNA sequencing approach in order to further refine the phylogeny of the Indian subcontinent specific segment of haplogroup R. In particular, we first added 35 novel complete mtDNA sequences from haplogroup R. In order to explore the correlations between genes, languages and geography in Indian subcontinent we carried out high resolution genotyping and phylogeographic detailed analyses on R7 – one of the intriguing haplogroups which shows particularly high frequency among the Austroasiatic (Munda) speaking groups of India.

After addition of 35 novel sequences, the newly reconstructed South Asian specific R tree revealed eight new subclades within six haplogroup R branches autochthonous for Indian subcontinent (supplementary Fig. 1. Ref. IV). We refine here internal topology of haplogroups R5-8, R30 and R31. Geographically, the spread of R7 in India is centered around the AA "heartland" (Bihar, Jharkhand, and Chhattisgarh). PC analysis based on frequency data of hg R subclades confirmed that majority of Munda speaking populations cluster separately from others mainly because of hg R7 (Fig. 2. Ref IV). Based on these preliminary results, we focused on R7 as a potential AA-associated marker. In general, the elevated frequency of hg R7 among the AA speakers of India can be explained by two alternative scenarios. Firstly, one may consider a possible origin of R7 among AA (Munda) speakers, possibly already outside India. Under this scenario the presence of R7 in some Dravidian and Indo-European speaking communities would be explained by its later introgression from the Munda communities or by language shift of some Munda speaking groups into Dravidian/Indo-European languages. Secondly, origin of R7 may lie among non-AA populations of India, with the presently observable higher frequency of R7 among AA resulting from founder effect(s) due to random genetic drift. To test these two scenarios, we carried out a detailed analysis of R7 mtDNAs in populations speaking different subgroups of AA languages, as well as among IE and Dravidian-speaking populations of Indian subcontinent.

Complete mtDNA sequence-based topology of hg R7 divulges two deep-rooted subclades (Fig. 1, Ref. IV). Interestingly all the AA individuals coalesce to the founder R7a1 that dates back to approximately 3 or 7 thousand YBP, depending on the mutation rate used. The coalescent times of R7 variation among Dravidians and Indo-Europeans are older. In other words, the only R7 lineage found by us in AA speakers of India *i.e.* R7a1 is nested within the R7 lineages found among Dravidian and Indo-European speakers of India. The higher diversity of R7a and R7b sub-clades among non Austroasiatic populations of India suggests that the source of haplogroup R7 is not among the maternal ancestors of all Austroasiatic tribal groups and that they have acquired this haplogroup via local admixture together with the rest of the South Asian mtDNA lineages that make up their extant maternal lineage pool. Furthermore, the presence of only a single recent founder branch of R7, *i.e.* R7a1, among widely dispersed AA populations of India supports the founder event scenario by introgression of this lineage from the local non-AA populations before the range expansion of Munda speaking populations within India. The occurrence of R7a1 among Dravidian and Indo-European speaking populations living close to the AA populations (Fig. 3. Ref IV), could however be explained by language shift or secondary admixture with AA speakers. Sub-haplogroup R7b appears to be restricted to Dravidian-speakers of the Southern part of India. Nonetheless, this haplogroup has also been reported in two Indo-European populations (Kolcha and Rathwa) whose local tradition, although, speaks about their ancient split from the Gond population of Central India and further migration to Gujarat. Thus, from the given data, it is most parsimonious to conjecture that R7 originated in India among non-AA, possibly in Dravidian speaking populations. The Mantel test suggests a significant and high correlation coefficient between genes and geography for hg R7 (0.299, $p<0.05$), strongest among the different R sub-clades observed in India, while none of the R sub-clades demonstrated significant correlation between gene and language (table 4 Ref IV). The spatial autocorrelation analysis favored a clinal pattern of the distribution of hg R7.

We also sought to examine whether the spread of R7 among the different Munda sub-groups in India, as defined by the language trees is uniform or not. This would be expected if R7 was present among the ancestral AA speakers prior to the diversification of the language family into numerous branches. Consistent with the non-AA origin of R7, we found the distribution of R7a1 among AA populations to be profoundly skewed towards the Kherwari sub-branch of the North Munda languages which accounts for ~90% of the AA R7 samples (Fig. 5 Ref IV). Accordingly, R7 is very rare in the South Munda group. It is completely absent in Koraput Munda speakers and marginally present only in the Kharia tribe of Madhya Pradesh (in total 3 out of 431 South Munda samples). This finding yet again fortifies the argument that only a subset of Indian AA groups has acquired one sublineage of R7a1 *in-situ* after their arrival to Indian subcontinent from local non-AA groups through admixture. Thus, we fail to find from the extant maternal lineage pool of the Austroasiatic

speakers of India any major lineages that show signs of potential origin outside India. Therefore, the enigma of the origin and demographic past of the AA speakers in India still stands there. While the East Asian contribution to their paternal gene pool seems evident, the maternal side of their genetic heritage appears to be autochthonous to Indian subcontinent.

5. CONCLUSIONS

- The maternal genepool of human population in India harbors several deep rooting lineages of macrohaplogroup M, suggesting *in-situ* origin of these clades in South Asia. The predominant autochthonous lineages of macrohaplogroup M in South Asia illustrate a non-nested starlike structure and don't share any significant early offshoot branches in the mtDNA tree with East Asian or Oceanian populations. The distribution of deeply rooted branches within haplogroup M across Asia and Oceania provides an additional evidence for a single migration route stemmed from Africa.
- Consistent with their geographical and cultural separation, our reanalysis of the Andamanese-specific lineage M31 divulged two clear-cut sub-population-specific branches structuring-M31a1 present in Onge and Jarwa individuals, while M31a2 was exclusively found amongst Great Andamanese.
- The infiltration of tribes in the caste system is in congruence with the social mobility or caste-tribe continuum. Thus, the contrasting geographic distribution of Scheduled Castes and Scheduled Tribes in South Asia can be observed. Our genetic analysis on Mushar population of North India reflects the cultural process in South Asia where language shift (replacement) without any large genetic exchange (*i.e.* cultural diffusion only) is a widespread phenomenon. This analysis offers a prime example of how language change is linked with the cultural diffusion in South Asia.
- The high resolution molecular study on haplogroup R7 revealed lack of deeply rooted phylogenetic structure and a clear-cut phylogeography in a more derived branch (*i.e.* R7a1) in Munda speakers. Therefore, the highest frequency but low diversity of haplogroup R7 among Munda speakers is by far more parsimoniously explained by their relatively recent regional admixture with local populations, followed by strong founder effect, rather than being a haplogroup, specific for Munda speakers, which has spread later on to Dravidians and Indo-Europeans.
- We draw a general conclusion, based on the age of the largely autochthonous South Asian matrilineal ancestry that is manifold older than any of the assumed language family, such as Indo-European, Dravidian or Austroasiatic, that the genetic landscape of the sub-continent is primarily explained by geography, not linguistics. Yet the latter has had its contribution, in particular at local levels, and together with impacts from social stratification, is very much worthwhile to consider and study in greater details in future.

REFERENCES

Abbi, A., 2006. Endangered Languages of the Andaman islands. Lincom Europa, Munich.

Abicht, A., Stucka, R., Karcagi, V., Herczegfalvi, A., Horváth, R., Mortier, W., Schara, U., Ramaekers, V., Jost, W., Brunner, J., Janssen, G., Seidel, U., Schlotter, B., Müller-Felber, W., Pongratz, D., Rüdel, R., Lochmüller, H., 1999. A common mutation (epsilon1267delG) in congenital myasthenic patients of Gypsy ethnic origin. Neurology 53, 1564–1569.

Abu-Amero, K.K., Larruga, J.M., Cabrera, V.M., González, A.M., 2008. Mitochondrial DNA structure in the Arabian Peninsula. BMC Evol Biol 8, 45.

Adcock, G.J., Dennis, E.S., Easteal, S., Huttley, G.A., Jermiin, L.S., Peacock, W.J., Thorne, A., 2001. Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins. Proc Natl Acad Sci USA 98, 537–542.

Agrawal, S., Khan, F., Pandey, A., Tripathi, M., Herrera, R., 2005. YAP, signature of an African-Middle Eastern migration into northern India. Curr Sci 88, 1977–1980.

Alam, S., Ali, M.E., Ferdous, A., Hossain, T., Hasan, M.M., Akhteruzzaman, S., 2010. Haplotype diversity of 17 Y-chromosomal STR loci in the Bangladeshi population. Forensic science international. Genetics 4, e59–e60.

Allchin, B., Allchin, F.R. (Ed.), 1997. Origins of a civilization. Viking Adult.

Al-Zahery, N., Semino, O., Benuzzi, G., Magri, C., Passarino, G., Torroni, A., Santachiara-Benerecetti, A.S., 2003. Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. Mol Phylogenet Evol 28, 458–72.

Amato, R., Pinelli, M., Monticelli, A., Marino, D., Miele, G., Cocozza, S., 2009. Genome-wide scan for signatures of human population differentiation and their relationship with natural selection, functional pathways and diseases. PloS one 4, e7927.

Ambrose, S.H., 2003. Did the super-eruption of Toba cause a human population bottleneck? Reply to Gathorne-Hardy and Harcourt-Smith. J Hum Evol 45, 231–237.

Ambrose, S., 2008. Small Things Remembered: Origins of Early Microlithic Industries in Sub-Saharan Africa. 12, 9–29.

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

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

Aquadro, C.F., Greenberg, B.D., 1983. Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics 103, 287–312.

Ashrama, A., 1944. Essentials of Hinduism. Advaita ashrama.

Athreya, S., 2007. Was Homo heidelbergensis in South Asia? A test using the Narmada fossil from central India, in: Petraglia, M.D., Allchin, B. (Ed.), The evolution and history of human populations in South Asia. Springer Verlag, pp. 464.

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. Mol Biol Evol 25, 468–474.

Auton, A., Bryc, K., Boyko, A.R., Lohmueller, K.E., Novembre, J., Reynolds, A., Indap, A., Wright, M.H., Degenhardt, J.D., Gutenkunst, R.N., King, K.S., Nelson, M.R., Bustamante, C.D., 2009. Global distribution of genomic diversity underscores rich complex history of continental human populations. Genome Res 19, 795–803.

Ayub, Q., Tyler-Smith, C., 2009. Genetic variation in South Asia: assessing the influences of geography, language and ethnicity for understanding history and disease risk. Brief Funct Genomic Proteomic 8, 395–404.

Baig, M.M., Khan, A.A., Kulkarni, K.M., 2004. Mitochondrial DNA diversity in tribal and caste groups of Maharashtra (India) and its implication on their genetic origins. Ann Hum Genet 68, 453–460.

Balgir, R.S., 2006. Do tribal communities show an inverse relationship between sickle cell disorders and glucose-6-phosphate dehydrogenase deficiency in malaria endemic areas of Central-Eastern India? Homo 57, 163–176.

Balgir, R.S., 2008. Hematological profile of twenty-nine tribal compound cases of hemoglobinopathies and G-6-PD deficiency in rural Orissa. Indian J Med Sci 62, 362–371.

Ballinger, S.W., Schurr, T.G., Torroni, A., Gan, Y.Y., Hodge, J.A., Hassan, K., Chen, K.H., Wallace, D.C., 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics 130, 139–152.

Bamshad, M., Fraley, A.E., Crawford, M.H., Cann, R.L., Busi, B.R., Naidu, J.M., Jorde, L.B., 1996. mtDNA variation in caste populations of Andhra Pradesh, India. Hum Biol 68, 1–28.

Bamshad, M.J., Watkins, W.S., Dixon, M.E., Jorde, L.B., Rao, B.B., Naidu, J.M., Prasad, B.V., Rasanayagam, A., Hammer, M.F., 1998. Female gene flow stratifies Hindu castes. Nature 395, 651–652.

Bamshad, M., Kivisild, T., Watkins, W.S., Dixon, M.E., Ricker, C.E., Rao, B.B., Naidu, J.M., Prasad, B.V., Reddy, P.G., Rasanayagam, A., Papiha, S.S., Villems, R., Redd, A.J., Hammer, M.F., Nguyen, S.V., Carroll, M.L., Batzer, M.A., Jorde, L.B., 2001. Genetic evidence on the origins of Indian caste populations. Genome Res 11, 994–1004.

Bandelt, H.J., Kong, Q.P., Richards, M., Macaulay, V., 2006. Estimation of Mutation Rates and Coalescence Times: Some Caveats, in: Bandelt, H.J., Macaulay, V., Richards, M., Bujnicki, J.M. (Ed.), Nucleic Acids and Molecular Biology. Springer-Verlag, Berlin Heidelberg, pp. 47–90.

Bandelt, H.J., Olivieri, A., Bravi, C., Yao, Y.G., Torroni, A., Salas, A., 2007a. 'Distorted' mitochondrial DNA sequences in schizophrenic patients. Eur J Hum Genet 15, 400–2; author reply 402.

Bandelt, H.J., Yao, Y.G., Salas, A., 2007b. The search of 'novel' mtDNA mutations in hypertrophic cardiomyopathy: MITOMAPping as a risk factor. Int J Cardiol

Bandyopadhyay, J., Perveen, S., 2002. The Interlinking of Indian Rivers:Some Questions on the Scientific, Economic and Environmental Dimensions of the Proposal, Interlinking Indian Rivers: Bane or Boon?, Kolkata, India

Banerjee, B., 1979. Hindu culture, custom, and ceremony. Agam.

Barik, S.S., Sahani, R., Prasad, B.V.R., Endicott, P., Metspalu, M., Sarkar, B.N., Bhattacharya, S., Annapoorna, P.C.H., Sreenath, J., Sun, D., Sanchez, J.J., Ho, S.Y.W., Chandrasekar, A., Rao, V.R., 2008. Detailed mtDNA genotypes permit a reassessment of the settlement and population structure of the Andaman Islands. Am J Phys Anthropol 136, 19–27.

Barnabas, S., Apte, R.V., Suresh, C.G., 1996. Ancestry and interrelationships of the Indians and their relationship with other world populations: a study based on mitochondrial DNA polymorphisms. Ann Hum Genet 60, 409–422.

Barnabas, S., Shouche, Y., Suresh, C.G., 2006. High-resolution mtDNA studies of the Indian population: implications for palaeolithic settlement of the Indian subcontinent. Ann Hum Genet 70, 42–58.

Bar-Yosef, O., Kuhn, S.L., 1999. The Big Deal about Blades: Laminar Technologies and Human Evolution. BMC Biol 101, 322–338.

Basu, A., Mukherjee, N., Roy, S., Sengupta, S., Banerjee, S., Chakraborty, M., Dey, B., Roy, M., Roy, B., Bhattacharyya, N.P., Roychoudhury, S., Majumder, P.P., 2003. Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res 13, 2277–2290.

Battersby, B.J., Loredo-Osti, J.C., Shoubridge, E.A., 2003. Nuclear genetic control of mitochondrial DNA segregation. Nat Genet 33, 183–186.

Behar, D.M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S., Comas, D., Bertranpetit, J., Quintana-Murci, L., Tyler-Smith, C., Wells, R.S., Rosset, S., 2008a. The Dawn of Human Matrilineal Diversity. The American Journal of Human Genetics 82, 1130–1140.

Behar, D.M., Metspalu, E., Kivisild, T., Rosset, S., Tzur, S., Hadid, Y., Yudkovsky, G., Rosengarten, D., Pereira, L., Amorim, A., Kutuev, I., Gurwitz, D., Bonne-Tamir, B., Villems, R., Skorecki, K., 2008b. Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. PLoS ONE 3, e2062.

Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., Khusnutdinova, E.K., Balanovsky, O., Semino, O., Pereira, L., Comas, D., Gurwitz, D., Bonne-Tamir, B., Parfitt, T., Hammer, M.F., Skorecki, K., Villems, R., 2010. The genome-wide structure of the Jewish people. Nature Jun 9.

Beja-Pereira, A., Luikart, G., England, P.R., Bradley, D.G., Jann, O.C., Bertorelle, G., Chamberlain, A.T., Nunes, T.P., Metodiev, S., Ferrand, N., Erhardt, G., 2003. Gene-culture coevolution between cattle milk protein genes and human lactase genes. Nat Genet 35, 311–313.

Beja-Pereira, A., Caramelli, D., Lalueza-Fox, C., Vernesi, C., Ferrand, N., Casoli, A., Goyache, F., Royo, L.J., Conti, S., Lari, M., Martini, A., Ouragh, L., Magid, A., Atash, A., Zsolnai, A., Boscato, P., Triantaphylidis, C., Ploumi, K., Sineo, L., Mallegni, F., Taberlet, P., Erhardt, G., Sampietro, L., Bertranpetit, J., Barbujani, G., Luikart, G., Bertorelle, G., 2006. The origin of European cattle: evidence from modern and ancient DNA. Proc Natl Acad Sci U S A 103, 8113–8118.

Bengston, J., 1996. “Nihali and Ainu.”. Mother Tongue 2, 51–55.

Bergstrom, C.T., Pritchard, J., 1998. Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. Genetics 149, 2135–2146.

Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J.A., Rhodes, M., Reich, D.E., Hirschhorn, J.N., 2004. Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74, 1111–1120.

Bhadra, B., Gupta, A., Sharma, J., 2009. Saraswati Nadi in Haryana and its linkage with the Vedic Saraswati River -Integrated study based on satellite images and ground based information. Journal of the Geological Society of India 73, 273–288.

Bhatia, H., Rao, V., 1986. Genetic Atlas of Indian Tribes. Institute of Immuno-haematology, New Delhi: Indian Council of Medical Research, 242–254.

Birky, C.W., Demko, C.A., Perlman, P.S., Strausberg, R., 1978. Uniparental inheritance of mitochondrial genes in yeast: dependence on input bias of mitochondrial DNA and preliminary investigations of the mechanism. Genetics 89, 615–651.
Bittles, A., 2003. Consanguineous marriage and childhood health. Dev Med Child Neurol 45, 571–576.
Bittles, A.H., Black, M.L., 2010. Evolution in health and medicine Sackler colloquium: Consanguinity, human evolution, and complex diseases. Proc Natl Acad Sci U S A 107 Suppl 1, 1779–1786.
Black, M.L., Dufall, K., Wise, C., Sullivan, S., Bittles, A.H., 2006. Genetic ancestries in northwest Cambodia. Ann Hum Biol 33, 620–627.
Blazek, V., Boisson, C., 1992. The diffusion of agricultural terms from Mesopotamia. Archiv Orientalni 60, 16–23.
Blazek, V., 1999. Elam, a bridge between Ancient Near East and Dravidian India, in: Blench, R. (Ed.), Archaeology and Language IV. One World Archaeology, London, pp. 48–78.
Blench, R., 2005. FROM THE MOUNTAINS TO THE VALLEYS, in: Sagart, L., Blench, R., Sanchez-Mazas, A. (Ed.), The peopling of East Asia. Routledge, pp. 323.
Blench, R., Dendo, M., 2007. The language of the Shom Pen: a language isolate in the Nicobar islands. Mother Tongue XII , 179–202.
Blench, R., 2008. Re-evaluating the linguistic prehistory of South Asia, in T. Osada, A. Uesugi (Ed.), Kyoto: Indus Project, Research Institute for Humanity and Nature.
Blench, R.M. 2007. The language of the Shom Pen: a language isolate in the Nicobar Islands. Mother Tongue XII, 179–202.
Blum, M., Rosenberg, N., 2007. Estimating the number of ancestral lineages using a maximum-likelihood method based on rejection sampling. Genetics 176, 1741.
Boivin, N., 2007. Anthropological, historical, archaeological and genetic perspectives on the origins of caste in South Asia. The Evolution and History of Human Populations in South Asia , 341–361.
Bosch, E., Calafell, F., Rosser, Z.H., Norby, S., Lynnerup, N., Hurles, M.E., Jobling, M.A., 2003. High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. Hum Genet 112, 353–63. Epub 2003 Feb 20.
Bouwer, S., Angelicheva, D., Chandler, D., Seeman, P., Tournev, I., Kalaydjieva, L., 2007. Carrier rates of the ancestral Indian W24X mutation in GJB2 in the general Gypsy population and individual subisolates. Genet Test 11, 455–458.
Breton, S., Beaupré, H.D., Stewart, D.T., Hoeh, W.R., Blier, P.U., 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends Genet 23, 465–474.
Brodd, J., Gregory, L., 2003. World Religions: A Voyage of Discovery St Marys Pr
Brown, W.M., George, M., Wilson, A.C., 1979. Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci U S A 76, 1967–1971.
Bryk, J., Hardouin, E., Pugach, I., Hughes, D., Strotmann, R., Stoneking, M., Myles, S., 2008. Positive selection in East Asians for an EDAR allele that enhances NF-kappaB activation. PLoS ONE 3, e2209.
Bryson, R., Swain, A., 1981. Holocene variations in monsoon rainfall in Rajasthan. Quaternary Research 16, 135–145.
Cann, R.L., Wilson, A.C., 1983. Length mutations in human mitochondrial DNA. Genetics 104, 699–711.
Cann, R.L., Brown, W.M., Wilson, A.C., 1984. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics 106, 479–499.

Cann, R.L., Stoneking, M., Wilson, A.C., 1987. Mitochondrial DNA and human evolution. Nature 325, 31–36.
Cann, R.L., 1994. mtDNA and Native Americans: a Southern perspective. Am J Hum Genet 55, 7–11.
Cann, R.L., 2001. Genetic clues to dispersal in human populations: retracing the past from the present. Science 291, 1742–178.
Caratini, C., Bentaleb, I., Fontugne, M., Morzadec-Kerfourn, M., Pascal, J., Tissot, C., 1994. A less humid climate since ca. 3500 yr B.P. from marine cores off Karwar, western India. Palaeogeography, Palaeoclimatology, Palaeoecology 109, 371–384.
Carvajal-Carmona LG, Soto JD, N, P., Ortziz-Barrientos D, C, D., Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, A, R.L., 2000. Strong Amerind/White Sex Bias and a Possible Sephardic Contribution among the Founders of a Population in Northwest Colombia. Am. J. Hum. Genet. 67, 1287–1295.
Carvalho-Silva, D.R., Santos, F.R., Rocha, J., Pena, S.D., 2001. The phylogeography of Brazilian Y-chromosome lineages. Am J Hum Genet 68, 281–286.
Carvalho-Silva, D., Tyler-Smith, C., 2008. The Grandest Genetic Experiment Ever Performed on Man?-A Y-Chromosomal Perspective on Genetic Variation in India. INTERNATIONAL JOURNAL OF HUMAN GENETICS 8, 21.
Cavalli-Sforza, L.L., Edwards, A.W., 1967. Phylogenetic analysis. Models and estimation procedures. Am J Hum Genet 19, 233–257.
Cavalli-Sforza, L.L., Piazza, A., Menozzi, P., Mountain, J., 1988. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc Natl Acad Sci U S A 85, 6002–6006.
Cavalli-Sforza, L.L., Menozzi, P., Piazza, A., 1994. The history and geography of human genes Princeton University Press, Princeton, N.J.
Cavalli-Sforza, L., 1996. The spread of agriculture and nomadic pastoralism: Insights from the genetics, linguistics and archaeology, in: Harris, D.R. (Ed.), The origins and spread of agriculture and pastoralism in Eurasia. Routledge, , pp. 594.
Chakravarti, A., 2009. Human genetics: Tracing India's invisible threads. Nature 461, 487–488.
Chandrasekar, A., Saheb, S.Y., Gangopadyaya, P., Gangopadyaya, S., Mukherjee, A., Basu, D., Lakshmi, G.R., Sahani, A.K., Das, B., Battacharya, S., Kumar, S., Xaviour, D., Sun, D., Rao, V.R., 2007. YAP insertion signature in South Asia. Ann Hum Biol 34, 582–586.
Chandrasekar, A., Kumar, S., Sreenath, J., Sarkar, B.N., Urade, B.P., Mallick, S., Bandopadhyay, S.S., Barua, P., Barik, S.S., Basu, D., Kiran, U., Gangopadhyay, P., Sahani, R., Prasad, B.V.R., Gangopadhyay, S., Lakshmi, G.R., Ravuri, R.R., Padmaja, K., Venugopal, P.N., Sharma, M.B., Rao, V.R., 2009. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. PloS one 4, e7447.
Charlesworth, B., 1996. The evolution of chromosomal sex determination and dosage compensation. Curr Biol 6, 149–62.
Charlesworth, B., Charlesworth, D., 2000. The degeneration of Y-chromosomes. Philos Trans R Soc Lond B Biol Sci 355, 1563–1572.
Charlesworth, D., Willis, J.H., 2009. The genetics of inbreeding depression. Nat Rev Genet 10, 783–796.
Chauhan, P., 2010. “Review of 'Petraglia, M.D. and B. Allchin (eds.) – The Evolution and History of Populations in South Asia: Inter-Disciplinary Studies in Archaeo-

logy, Biological Anthropology, Linguistics and Genetics.'". PaleoAnthropology 2010, 64–71.
Chen, S., Lin, B.Z., Baig, M., Mitra, B., Lopes, R.J., Santos, A.M., Magee, D.A., Azevedo, M., Tarroso, P., Sasazaki, S., Ostrowski, S., Mahgoub, O., Chaudhuri, T.K., Zhang, Y.P., Costa, V., Royo, L.J., Goyache, F., Luikart, G., Boivin, N., Fuller, D.Q., Mannen, H., Bradley, D.G., Beja-Pereira, A., 2010. Zebu cattle are an exclusive legacy of the South Asia neolithic. Mol Biol Evol 27, 1–6.
Chinnery, P., 2006. Mitochondrial DNA in Homo Sapiens, in: Bandelt, H.J., Macaulay, V., Richards, D.M. (Ed.), Human mitochondrial DNA and the evolution of Homo sapiens. Springer Verlag, , pp. 271.
Clark, V.J., Sivendren, S., Saha, N., Bentley, G.R., Aunger, R., Sirajuddin, S.M., Stoneking, M., 2000. The 9-bp deletion between the mitochondrial lysine tRNA and COII genes in tribal populations of India. Hum Biol 72, 273–285.
Clarkson, C., Petraglia, M., Korisettar, R., Haslam, M., Boivin, N., Crowther, A., Ditchfield, P., Fullar, D., Miracle, P., Harris, C., Connell, K, James, H., Koshy, J., 2009. The oldest and longest enduring microlithic sequence in India: 35,000 years of modern human occupation and change at the Jwalapuram Locality 9 rockshelter. Antiquity 83, 326–348.
Cooper, Z., 2002. Archaeology and History: Early settlements in the Andaman Islands Oxford University press, New Delhi and Oxford.
Cordaux, R., Saha, N., Bentley, G., Aunger, R., Sirajuddin, S., Stoneking, M., 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. Eur J Hum Genet 3, 253–264.
Cordaux, R., Deepa, E., Vishwanathan, H., Stoneking, M., 2004a. Genetic evidence for the demic diffusion of agriculture to India. Science 304, 1125.
Cordaux, R., Weiss, G., Saha, N., Stoneking, M., 2004b. The northeast Indian passageway: a barrier or corridor for human migrations? Mol Biol Evol 21, 1525–1533.
Cordaux, R., Aunger, R., Bentley, G., Nasidze, I., Sirajuddin, S.M., Stoneking, M., 2004c. Independent origins of Indian caste and tribal paternal lineages. Curr Biol 14, 231–235.
Costantini, L., 1984. The beginning of agriculture of the Kachi plain: the evidence of Mehrgarh, in: Allchin, B. (Ed.), South Asian Archaeology 1981. UK, pp. 346.
Coyne, J.A., Hoekstra, H.E., 2007. Evolution of protein expression: new genes for a new diet. Curr Biol 17, R1014-R1016.
Currat, M., Excoffier, L., 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. PLoS Biol 2, e421.
deMenocal, P.B., 1995. Plio-Pleistocene African climate. Science 270, 53–59.
deMenocal, P.B., 2001. Cultural responses to climate change during the late Holocene. Science 292, 667–673.
Dennell, R., Rendell, H., M, H., Moth, E., 1992. A 45,000-Year-Old Open-air Paleolithic Site at Riwat, Northern Pakistan. Journal of Field Archaeology 19
Dennell, R., Roebroeks, W., 2005. An Asian perspective on early human dispersal from Africa. Nature 438, 1099–1104.
Dennell, R., 2007. "Resource-rich, stone-poor": Early hominin land use in large river systems of northern India and Pakistan, in: Petraglia, M.D., Allchin, B. (Ed.), The evolution and history of human populations in South Asia. Springer Verlag, pp. 464.
Deraniyagala, S.U., 1984. Mesolithic stone tool technology at 28,000 B.P. in Sri Lanka. Ancient Ceylon 5, 105–108.

Diamond, J., 1997. Guns, Germs and Steel: The Fates of Human Societies Jonathan Cape, London.
Diamond, J., 2002. Evolution, consequences and future of plant and animal domestication. Nature 418, 700–707.
Diamond, J., Bellwood, P., 2003. Farmers and their languages: the first expansions. Science 300, 597–603.
Diffloth, G., 2005. The contribution of linguistic palaeontology to the homeland of Austroasiatic, in: Sagart, L., Blench, R., Sanchez-Mazas, A. (Ed.), The peopling of East Asia. Routledge, , pp. 323.
Diffloth, Gérard. 2009. More on Dvaravati Old Mon. Paper presented at the Fourth International Conference on Austroasiatic Linguistics, Mahidol University at Salaya, 29 October 2009.
Dungdung, G., 2003. Human Rights: Wrong with Musahars, http://jharkhandmirror.org/2009/07/25/human-rights-wrong-with-musahars.
Dunstan, W., 1998. The ancient Near East, Elenchus of Biblica. , pp. 708.
Eaaswarkhanth, M., Haque, I., Ravesh, Z., Romero, I.G., Meganathan, P.R., Dubey, B., Khan, F.A., Chaubey, G., Kivisild, T., Tyler-Smith, C., Singh, L., Thangaraj, K., 2010. Traces of sub-Saharan and Middle Eastern lineages in Indian Muslim populations. Eur J Hum Genet 18, 354–63.
Eberhart, C.G., Maines, J.Z., Wasserman, S.A., 1996. Meiotic cell cycle requirement for a fly homologue of human Deleted in Azoospermia. Nature 381, 783–785.
Elson, J.L., Turnbull, D.M., Howell, N., 2004. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. Am J Hum Genet 74, 229–238.
Endicott, P., Gilbert, M.T.P., Stringer, C., Lalueza-Fox, C., Willerslev, E., Hansen, A.J., Cooper, A., 2003a. The genetic origins of the Andaman Islanders. Am J Hum Genet 72, 178–184.
Endicott, P., Macaulay, V., Kivisild, T., Stringer, C., Cooper, A., 2003b. Reply to Cordaux and Stoneking:. Am J Hum Genet 72, 1590–1593.
Endicott, P., Metspalu, M., Stringer, C., Macaulay, V., Cooper, A., Sanchez, J.J., 2006. Multiplexed SNP typing of ancient DNA clarifies the origin of andaman mtDNA haplogroups amongst south Asian tribal populations. PLoS ONE 1, e81.
Endicott, P., Metspalu, M., Kivisild, T., 2007. Genetic evidence on modern human dispersals in South Asia: Y-chromosome and mitochondrial DNA perspectives, in: Petraglia, M., Allchin, B. (Ed.), The Evolution and History of Human Populations in South Asia: Inter-disciplinary Studies in Archaeology, Biological Anthropology, Linguistics and Genetics. Springer/Kluwer Academic Publishers, pp. 390.
Endicott, P., Ho, S.Y., 2008. A Bayesian evaluation of human mitochondrial substitution rates. Am J Hum Genet 82, 895–902.
Endicott, P., Ho, S.Y., Metspalu, M., Stringer, C., 2009. Evaluating the mitochondrial timescale of human evolution. Trends Ecol Evol 24, 515–521.
Eshed, V., Gopher, A., Hershkovitz, I., 2006. Tooth wear and dental pathology at the advent of agriculture: new evidence from the Levant. Am J Phys Anthropol 130, 145–159.
Eswaran, V., Harpending, H., Rogers, A.R., 2005. Genomics refutes an exclusively African origin of humans. J Hum Evol 49, 1–18.
Fagundes, N.J., Kanitz, R., Eckert, R., Valls, A.C., Bogo, M.R., Salzano, F.M., Smith, D.G., Silva, W.A., Zago, M.A., Ribeiro-dos-Santos, A.K., Santos, S.E., Petzl-Erler, M.L., Bonatto, S.L., 2008. Mitochondrial population genomics supports a single pre-

Clovis origin with a coastal route for the peopling of the Americas. Am J Hum Genet 82, 583–592.
Farmer, S., Sproat, R., Witzel, M., 2004. The Collapse of the Indus-Script Thesis: The Myth of a Literate Harappan. Electronic Journal of Vedic Studies, 11.
Ferdous, A., Ali, M.E., Alam, S., Hasan, M., Hossain, T., Akhteruzzaman, S., 2009. Forensic evaluation of STR data for the PowerPlex 16 System loci in a Bangladeshi population. Leg Med (Tokyo) 11, 198–199.
Fernandes, S., Huellen, K., Goncalves, J., Dukal, H., Zeisler, J., Rajpert De Meyts, E., Skakkebaek, N.E., Habermann, B., Krause, W., Sousa, M., Barros, A., Vogt, P.H., 2002. High frequency of DAZ1/DAZ2 gene deletions in patients with severe oligozoospermia. Mol Hum Reprod 8, 286–98.
Fernandes, S., Paracchini, S., Meyer, L.H., Floridia, G., Tyler-Smith, C., Vogt, P.H., 2004. A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in Y haplogroup N. Am J Hum Genet 74, 180–187.
Fernandes, A.T., Fernandes, S., Goncalves, R., Sa, R., Costa, P., Rosa, A., Ferras, C., Sousa, M., Brehm, A., Barros, A., 2006. DAZ gene copies: evidence of Y-chromosome evolution. Mol Hum Reprod 12, 519–523.
Feuerstein, G., Kak, S., Frawley, D. (Ed.), 2001. In Search of the Cradle of Civilization Quest Books (IL)
Field, J., Lahr, M., 2005. Assessment of the Southern Dispersal: GIS-Based Analyses of Potential Routes at Oxygen Isotopic Stage 4. Journal of World Prehistory 19, 1–45.
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, 88–108.
Fleitmann, D., Burns, S.J., Mudelsee, M., Neff, U., Kramers, J., Mangini, A., Matter, A., 2003. Holocene forcing of the Indian monsoon recorded in a stalagmite from southern Oman. Science 300, 1737–1739.
Fornarino, S., Pala, M., Battaglia, V., Maranta, R., Achilli, A., Modiano, G., Torroni, A., Semino, O., Santachiara-Benerecetti, S.A., 2009. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. BMC Evol Biol 9, 154.
Forster, P., Harding, R., Torroni, A., Bandelt, H.J., 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59, 935–945.
Forster, P., Matsumura, S., 2005. Evolution. Did early humans go north or south? Science 308, 965–966.
Fraser, A., 1992. The Gypsies Blackwell Publishers Ltd, Oxford.
Friedlaender, J., Schurr, T., Gentz, F., Koki, G., Friedlaender, F., Horvat, G., Babb, P., Cerchio, S., Kaestle, F., Schanfield, M., Deka, R., Yanagihara, R., Merriwether, D.A., 2005. Expanding Southwest Pacific mitochondrial haplogroups P and Q. Mol Biol Evol 22, 1506–1517.
Friedlaender, J.S., Friedlaender, F.R., Hodgson, J.A., Stoltz, M., Koki, G., Horvat, G., Zhadanov, S., Schurr, T.G., Merriwether, D.A., 2007. Melanesian mtDNA complexity. PLoS ONE 2, e248.
Fuller, D., 2003. An Agricultural Perspective on Dravidian Historical Linguistics: Archaeological Crop Packages, Livestock and Dravidian Crop Vocabulary, in: Bellwood, P., Renfrew, C. (Ed.), Examining the farming/language dispersal hypothesis. The McDonald Institute for Archaeological Research, Cambridge.
Fuller, D., Korisettar, R., Vankatasubbaiah, P., Jones, M., 2004. Early plant domestications in southern India: some preliminary archaeobotanical results. Vegetation History Archaeobotany 13, 115–129.

Fuller, D., 2006. Agricultural Origins and Frontiers in South Asia: A Working Synthesis. J World Prehist 20, 1–86.
Fuller, D., 2007. Non-Human Genetics, Agricultural Origins and Historical Linguistics in South Asia, in: Petraglia, M., Allchin, B. (Ed.), Vertebrate Paleobiology and Paleoanthropology. Springer, pp. 393–443.
Gadgil, M., 1997. Peopling of India, in: Rao, B.A.N.A. (Ed.), The Indian Human Heritage. Universities Press, Hyderabad, India, pp. 100–129.
Gaikwad, S., Kashyap, V., Brodsky, A., Meyer, C., Swinburne, I., Hall, G., Keenan, B., Liu, X., Fox, E., Silver, P., 2005. Molecular insight into the genesis of ranked caste populations of western India based upon polymorphisms across non-recombinant and recombinant regions in genome. Genome Biol 6, P10.
Gangal, K., Vahia, M.N., Adhikari, R., 2010. Spatio-temporal analysis of the Indus urbanization. Current Science 98, 846–852.
Gardner, M., Williamson, S., Casals, F., Bosch, E., Navarro, A., Calafell, F., Bertranpetit, J., Comas, D., 2007. Extreme individual marker FST values do not imply population-specific selection in humans: the NRG1 example. Hum Genet 121, 759–762.
Garris, A., Tai, T., Coburn, J., Kresovich, S., McCouch, S., 2005. Genetic structure and diversity in Oryza sativa L. Genetics 169, 1631.
Gaur, A., Vora, K., 1999. Ancient shorelines of Gujarat, India, during the Indus civilization (Late–Mid–Holocene): A study based on archaeological evidences. Curr Sci 77, 180–185.
Gayden, T., Cadenas, A.M., Regueiro, M., Singh, N.B., Zhivotovsky, L.A., Underhill, P.A., Cavalli-Sforza, L.L., Herrera, R.J., 2007. The Himalayas as a directional barrier to gene flow. Am J Hum Genet 80, 884–894.
Gilbert, M., Bandelt, H., Hofreiter, M., Barnes, I., 2005. Assessing ancient DNA studies. Trends in Ecology & Evolution 20, 541–544.
Giles, R.E., Blanc, H., Cann, H.M., Wallace, D.C., 1980. Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci U S A 77, 6715–6719.
Gimbutas, M., 1970. Proto-Indo-European Culture: The Kurgan Culture during the fifth, fourth, and third millenia B.C, in: Cardona, G., Hoenigswald, H.M., Senn, A. (Ed.), Indo-European and Indo-Europeans. University of Pennsylvania Press, Philadelphia, pp. 155–195.
Gita Press, http://www.gitapress.org/
Glacier, Y., 2006. The Homeland of the Early Rigvedic Rishis: The Saraswati Basin in Haryana, Vedic Culture and Its Continuity: Proceedings of National Seminar, Pratibha Prakashan, , pp. 47.
Glaszmann, J., 1987. Isozymes and classification of Asian rice varieties. TAG Theoretical and Applied Genetics 74, 21–30.
Glazko, G.V., Nei, M., 2003. Estimation of divergence times for major lineages of primate species. Mol Biol Evol 20, 424–434.
Goff, S., Ricke, D., Lan, T., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., 2002. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296, 92.
Good, I., Kenoyer, J., Meadow, R., 2008. New Evidence for Early Silk in the Indus Civilization, Nature Precedings.
Goodbred, S., and Kuehl, S., 2000. Enormous Ganges–Brahmaputra sediment discharge during strengthened early Holocene monsoon. Geology 28, 1083–1086.
Goodman, M., Porter, C.A., Czelusniak, J., Page, S.L., Schneider, H., Shoshani, J., Gunnell, G., Groves, C.P., 1998. Toward a phylogenetic classification of Primates

based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9, 585–598.
Goyandka, J., 1969. r mad Bhagavadg t: with Sanskrit text and English translation Gita Press
Gray, R.D., Atkinson, Q.D., 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature 426, 435–49.
Green, R.E., Krause, J., Ptak, S.E., Briggs, A.W., Ronan, M.T., Simons, J.F., Du, L., Egholm, M., Rothberg, J.M., Paunovic, M., Pääbo, S., 2006. Analysis of one million base pairs of Neanderthal DNA. Nature 444, 330–336.
Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.Y., Hansen, N.F., Durand, E.Y., Malaspinas, A.S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prufer, K., Meyer, M., Burbano, H.A., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A., Hober, B., Hoffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V.B., Golovanova, L.V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R.W., Johnson, P.L.F., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Paabo, S., 2010. A Draft Sequence of the Neandertal Genome. Science 328, 710–722.
Greenberg, B., Newbold, J., Sugino, A., 1983. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21, 33–49.
Gresham, D., Morar, B., Underhill, P.A., Passarino, G., Lin, A.A., Wise, C., Angelicheva, D., Calafell, F., Oefner, P.J., Shen, P., Tournev, I., de Pablo, R., Kučinskas, V., Perez-Lezaun, A., Marushiakova, E., Popov, V., Kalaydjieva, L., 2001. Origins and divergence of the Roma (gypsies). Am J Hum Genet 69, 1314–1331.
Gupta, A., Anderson, D., Pandey, D., and Singhvi, A., 2006. Adaptation and human migration, and evidence of agriculture coincident with changes in the Indian summer monsoon during the Holocene. Curr Sci 60, 1082–1090.
Gupta, A.K., Anderson, D.M., Overpeck, J.T., 2003. Abrupt changes in the Asian southwest monsoon during the Holocene and their links to the North Atlantic Ocean. Nature 421, 354–357.
Gurney, O.R. (Ed.), 1990. The Hittites Penguin Group USA.
Gutala, R., Carvalho-Silva, D.R., Jin, L., Yngvadottir, B., Avadhanula, V., Nanne, K., Singh, L., Chakraborty, R., Tyler-Smith, C., 2006. A shared Y-chromosomal heritage between Muslims and Hindus in India. Hum Genet 120, 543–551.
Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M., Charlesworth, B., Keightley, P.D., 2008. Direct estimation of the mitochondrial DNA mutation rate in Drosophila melanogaster. PLoS Biol 6, e204.
Hammer, M.F., Spurdle, A.B., Karafet, T., Bonner, M.R., Wood, E.T., Novelletto, A., Malaspina, P., Mitchell, R.J., Horai, S., Jenkins, T., Zegura, S.L., 1997. The geographic distribution of human Y-chromosome variation. Genetics 145, 787–805.
Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A.R., Zegura, S.L., 1998. Out of Africa and back again: nested cladistic analysis of human Y-chromosome variation. Mol Biol Evol 15, 427–441.
Hammer, M.F., Karafet, T.M., Park, H., Omoto, K., Harihara, S., Stoneking, M., Horai, S., 2006. Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y-chromosomes. J Hum Genet 51, 47–58.

Harihara, S., Saitou, N., Hirai, M., Gojobori, T., Park, K.S., Misawa, S., Ellepola, S.B., Ishida, T., Omoto, K., 1988. Mitochondrial DNA polymorphism among five Asian populations. Am J Hum Genet 43, 134–143.

Harihara, S., Hirai, M., Suutou, Y., Shimizu, K., Omoto, K., 1992. Frequency of a 9-bp deletion in the mitochondrial DNA among Asian populations. Hum Biol 64, 161–166.

Harpending, H., Eswaran, V., 2005. Tracing modern human origins. Science 309, 1995–7; author reply 1995.

Hatcher, B., 1994. The Cosmos is One Family'(Vasudhaiva Kutumbakam). Contributions to Indian Sociology.

Heath, S.C., Gut, I.G., Brennan, P., McKay, J.D., Bencko, V., Fabianova, E., Foretova, L., Georges, M., Janout, V., Kabesch, M., Krokan, H.E., Elvestad, M.B., Lissowska, J., Mates, D., Rudnai, P., Skorpen, F., Schreiber, S., Soria, J.M., Syvanen, A.C., Meneton, P., Hercberg, S., Galan, P., Szeszenia-Dabrowska, N., Zaridze, D., Genin, E., Cardon, L.R., Lathrop, M., 2008. Investigation of the fine structure of European populations with applications to disease association studies. Eur J Hum Genet 16, 1413–1429.

Hertzberg, M., Mickleson, K.N., Serjeantson, S.W., Prior, J.F., Trent, R.J., 1989. An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. Am J Hum Genet 44, 504–510.

Higham, C. (Ed.), 1996. The Bronze Age of Southeast Asia Cambridge Univ Pr

Higham, C., 2003. Languages and Farming Dispersals: Austroasiatic Languages and Rice Cultivation, in: Bellwood, P., Renfrew, C. (Ed.), Examining the farming/ language dispersal hypothesis. The McDonald Institute for Archaeological Research, Cambridge.

Ho, S.Y.W., Phillips, M.J., Cooper, A., Drummond, A.J., 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol 22, 1561–1568.

Ho, S.Y.W., Gilbert, M.T.P., 2010. Ancient mitogenomics. Mitochondrion 10, 1–11.

Hofer, T., Ray, N., Wegmann, D., Excoffier, L., 2009. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. Ann Hum Genet 73, 95–108.

Hofreiter, M., Serre, D., Poinar, H., Kuch, M., Pääbo, S., 2001. Ancient DNA. Nature Reviews Genetics 2, 353–359.

Hong, Y., Hong, B., Lin, Q., Zhu, Y., Shibata, Y., Hirota, M., Uchida, M., Leng, X., Jiang, H., Xu, H., Wang, H., Yi, L., 2003. Correlation between Indian Ocean summer monsoon and North Atlantic climate during the Holocene. Earth Planet. Sci. Lett. 211.

Howells, W.W. (Ed.), 1974. The Pacific islanders Scribner Book Company.

Hudjashov, G., Kivisild, T., Underhill, P.A., Endicott, P., Sanchez, J.J., Lin, A.A., Shen, P., Oefner, P., Renfrew, C., Villems, R., Forster, P., 2007. Revealing the prehistoric settlement of Australia by Y-chromosome and mtDNA analysis. Proc Natl Acad Sci U S A 104, 8726–8730.

Hughes, J.F., Skaletsky, H., Pyntikova, T., Graves, T.A., van Daalen, S.K.M., Minx, P.J., Fulton, R.S., McGrath, S.D., Locke, D.P., Friedman, C., Trask, B.J., Mardis, E.R., Warren, W.C., Repping, S., Rozen, S., Wilson, R.K., Page, D.C., 2010. Chimpanzee and human Y-chromosomes are remarkably divergent in structure and gene content. Nature 463, 536–539.

HUGO Pan-Asian SNP Consortium, Abdulla, M.A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S.K., Calacal, G.C., Chaurasia, A., Chen, C.H., Chen, J.,

Chen, Y.T., Chu, J., Cutiongco-de la Paz, E.M.C., De Ungria, M.C.A., Delfin, F.C., Edo, J., Fuchareon, S., Ghang, H., Gojobori, T., Han, J., Ho, S.F., Hoh, B.P., Huang, W., Inoko, H., Jha, P., Jinam, T.A., Jin, L., Jung, J., Kangwanpong, D., Kampuansai, J., Kennedy, G.C., Khurana, P., Kim, H.L., Kim, K., Kim, S., Kim, W.Y., Kimm, K., Kimura, R., Koike, T., Kulawonganunchai, S., Kumar, V., Lai, P.S., Lee, J.Y., Lee, S., Liu, E.T., Majumder, P.P., Mandapati, K.K., Marzuki, S., Mitchell, W., Mukerji, M., Naritomi, K., Ngamphiw, C., Niikawa, N., Nishida, N., Oh, B., Oh, S., Ohashi, J., Oka, A., Ong, R., Padilla, C.D., Palittapongarnpim, P., Perdigon, H.B., Phipps, M.E., Png, E., Sakaki, Y., Salvador, J.M., Sandraling, Y., Scaria, V., Seielstad, M., Sidek, M.R., Sinha, A., Srikummool, M., Sudoyo, H., Sugano, S., Suryadi, H., Suzuki, Y., Tabbada, K.A., Tan, A., Tokunaga, K., Tongsima, S., Villamor, L.P., Wang, E., Wang, Y., Wang, H., Wu, J.Y., Xiao, H., Xu, S., Yang, J.O., Shugart, Y.Y., Yoo, H.S., Yuan, W., Zhao, G., Zilfalil, B.A., Indian Genome Variation Consortium, 2009. Mapping human genetic diversity in Asia. Science 326, 1541–1545.

Hurles, M.E., Irven, C., Nicholson, J., Taylor, P.G., Santos, F.R., Loughlin, J., Jobling, M.A., Sykes, B.C., 1998. European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. Am J Hum Genet 63, 1793–1806.

Hurst, L.D., 2009. Evolutionary genomics: A positive becomes a negative. Nature 457, 543–544.

Indian Genome Variation Consortium, 2008. Genetic landscape of the people of India: a canvas for disease gene exploration. J Genet 87, 3–20.

Ingman, M., Kaessmann, H., Pääbo, S., Gyllensten, U., 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408, 708–713.

Ingman, M., Gyllensten, U., 2001. Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. J Hered 92, 454–61.

International HapMap Consortium, 2005. A haplotype map of the human genome. Nature 437, 1299–1320.

International HapMap Consortium, Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M.Y., Tsui, S.K.W., Xue, H., Wong, J.T.F., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallée, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., Song, Y.Q., Tam, P.K.H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851–861.

International Rice Genome Sequencing Project, 2005. The map-based sequence of the rice genome. Nature 436, 793–800.

Iyengar, R., 2009. Monsoon rainfall cycles as depicted in ancient Sanskrit texts. Curr Sci 97, 7.

Izagirre, N., Garcia, I., Junquera, C., de la Rua, C., Alonso, S., 2006. A Scan for Signatures of Positive Selection in Candidate Loci for Skin Pigmentation in Humans 10.1093/molbev/msl030. Mol Biol Evol 23, 1697–1706.

Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.C., Szpiech, Z.A., Degnan, J.H., Wang, K., Guerreiro, R., Bras, J.M., Schymick, J.C., Hernandez, D.G., Traynor, B.J., Simon-Sanchez, J., Matarin, M., Britton, A., van de Leemput, J., Rafferty, I., Bucan, M., Cann, H.M., Hardy, J.A., Rosenberg, N.A., Singleton, A.B., 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451, 998–1003.

James, H., Petraglia, M., 2005. Modern human origins and the evolution of behavior in the Late Pleistocene record of South Asia. Curr Anthropol 46, S3-S27.

Jarrige, J.F., 1981. Chronology of the earlier periods of the Greater Indus as seen from Mehrgarh, Pakistan, in: Allchin, B. (Ed.), South Asian Archaeology. Cambridge University Press, Cambridge, , pp. 21–28.

Jin, J., Huang, W., Gao, J.P., Yang, J., Shi, M., Zhu, M.Z., Luo, D., Lin, H.X., 2008. Genetic control of rice plant architecture under domestication. Nat Genet 40, 1365–1369.

Jobling, M.A., Tyler-Smith, C., 1995. Fathers and sons: the Y-chromosome and human evolution. Trends Genet 11, 449–456.

Jobling, M.A., Tyler-Smith, C., 2003. The human Y-chromosome: an evolutionary marker comes of age. Nat Rev Genet 4, 598–612.

Jobling, M., Hurles, M., Tyler-Smith, C., 2004. Human evolutionary genetics Garland Science, New York and Abingdon, United Kingdom.

Jorde, L.B., Bamshad, M., Rogers, A.R., 1998. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. Bioessays 20, 126–136.

Kalaydjieva, L., Calafell, F., Jobling, M.A., Angelicheva, D., de Knijff, P., Rosser, Z.H., Hurles, M.E., Underhill, P., Tournev, I., Marushiakova, E., Popov, V., 2001. Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y-chromosome and mitochondrial DNA lineages. Eur J Hum Genet 9, 97–104.

Kalaydjieva, L., Morar, B., Chaix, R., Tang, H., 2005. A newly discovered founder population: the Roma/Gypsies. Bioessays 27, 1084–1094.

Kalyanaraman, S., Sarasvati Research and Education Trust (Madras, India) (Ed.), 2008. Vedic River Sarasvati and Hindu civilization.

Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., Hammer, M.F., 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18, 830–8.

Karve, I., Malhotra, K., 1968. A biological comparison of eight endogamous groups of the same rank. Curr Anthropol 9, 109.

Kashyap, V.K., Guha, S., Sitalaximi, T., Bindu, G.H., Hasnain, S.E., Trivedi, R., 2006. Genetic structure of Indian populations based on fifteen autosomal microsatellite loci. BMC Genet 7, 28.

Kayser, M., Krawczak, M., Excoffier, L., Dieltjes, P., Corach, D., Pascali, V., Gehrig, C., Bernini, L.F., Jespersen, J., Bakker, E., Roewer, L., de Knijff, P., 2001a. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet 68, 990–1018.

Kayser, M., Brauer, S., Weiss, G., Schiefenhovel, W., Underhill, P.A., Stoneking, M., 2001b. Independent histories of human Y-chromosomes from Melanesia and Australia. Am J Hum Genet 68, 173–190.

Kayser, M., Brauer, S., Weiss, G., Schiefenhovel, W., Underhill, P., Shen, P., Oefner, P., Tommaseo-Ponzetta, M., Stoneking, M., 2003. Reduced Y-chromosome, but not

mitochondrial DNA, diversity in human populations from West New Guinea. Am J Hum Genet 72, 281–302.
Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovsky, L.A., Moyse-Faurie, C., Rutledge, R.B., Schiefenhoevel, W., Gil, D., Lin, A.A., Underhill, P.A., Oefner, P.J., Trent, R.J., Stoneking, M., 2006. Melanesian and Asian origins of Polynesians: mtDNA and Y-chromosome gradients across the Pacific. Mol Biol Evol 23, 2234–2244.
Kazanas, N., 1999. The R. gveda and Indo-Europeans. Annals of the Bhandarkar Oriental Research Institute.
Kennedy, K., 2001. Middle and Late Pleistocene Hominids of South Asia, in: Tobias, P., Rath, M., Moggi-Cecchi, J., Doyle, G. (Ed.), Humanity from African Naissance to Coming Millennia. Firenze University Press, Florence, pp. 167–174.
Kennedy, K.A.R., 2000. God Apes and Fossil Men: Palaeoanthropology in South Asia University of Michigan Press, Ann Arbor.
Kenoyer, J.M. (Ed.), 1998. Ancient cities of the Indus Valley civilization, Oxford University Press.
Khush, G., 1997. Origin, dispersal, cultivation and variation of rice. Plant Mol Biol 35, 25–34.
Kivisild, T., Bamshad, M.J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., Laos, S., Parik, J., Watkins, W.S., Dixon, M.E., Papiha, S.S., Mastana, S.S., Mir, M.R., Ferak, V., Villems, R., 1999a. Deep common ancestry of indian and western-Eurasian mitochondrial DNA lineages. Curr Biol 9, 1331–1334.
Kivisild, T., Kaldma, K., Metspalu, M., Parik, J., Papiha, S., Villems, R., 1999b. The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the Old World, in: Papiha, S., Deka, R., Chakraborty, R. (Ed.), Genomic diversity. Kluwer Academic/Plenum Publishers, pp. 135–152.
Kivisild, T., 2000a. PhD Thesis: The Origins of Southern and Western Eurasian Populations: an mtDNA Study, Tartu University, Departement of Evolutionary Biology, Institute of Cell and Molecular Biology, Tartu.
Kivisild, T., Papiha, S.S., Rootsi, S., Parik, J., Kaldma, K., Reidla, M., Laos, S., Metspalu, M., Pielberg, G., Adojaan, M., Metspalu, E., Mastana, S.S., Wang, Y., Gölge, M., Demirtas, H., Schnekenberg, E., Stefano, G.F., Geberhiwot, T., Claustres, M., Villems, R., 2000b. An Indian ancestry: a key for understanding human diversity in Europe and beyond, in: Renfrew, C., Boyle, K. (Ed.), Archaeogenetics: DNA and the population prehistory of Europe. McDonald Institute for Archaeological Research University of Cambridge, Cambridge, pp. 267–279.
Kivisild, T., Tolk, H.V., Parik, J., Wang, Y., Papiha, S.S., Bandelt, H.J., Villems, R., 2002. The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol 19, 1737–1751.
Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H.V., Stepanov, V., Golge, M., Usanga, E., Papiha, S.S., Cinnioglu, C., King, R., Cavalli-Sforza, L., Underhill, P.A., Villems, R., 2003a. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet 72, 313–332.
Kivisild, T., Rootsi, S., Metspalu, M., Metspalu, E., Parik, J., Kaldma, K., Usanga, E., Mastana, S., Papiha, S., Villems, R., 2003b. The genetics of language and farming spread in India, in: Bellwood, P., Renfrew, C. (Ed.), Examining the farming/language dispersal hypothesis. The McDonald Institute for Archaeological Research, Cambridge, pp. 215–222.

Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E., Parik, J., Geberhiwot, T., Usanga, E., Villems, R., 2004. Ethiopian mitochondrial DNA heritage: tracking gene flows across and around the Strait of Tears. Am J Hum Genet 75, 752–770.

Kivisild, T., Shen, P., Wall, D.P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P.A., Scharfe, C., Torroni, A., Scozzari, R., Modiano, D., Coppa, A., de Knijff, P., Feldman, M., Cavalli-Sforza, L.L., Oefner, P.J., 2006a. The role of selection in the evolution of human mitochondrial genomes. Genetics 172, 373–387.

Kivisild, T., Metspalu, M., Bandelt, H.J., Richards, M., Villems, R., 2006b. The world mtDNA phylogeny, in: Bandelt, H.J., Macaulay, V., Richards, M. (Ed.), Human mitochondrial DNA and the evolution of Homo sapiens. Springer-Verlag, Heidelberg.

Klarić, I.M., Salihović, M.P., Lauc, L.B., Zhivotovsky, L.A., Rootsi, S., Janićijević, B., 2009. Dissecting the molecular architecture and origin of Bayash Romani patri-lineages: genetic influences from South-Asia and the Balkans. Am J Phys Anthropol 138, 333–342.

Kondrashov, A.S., 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Hum Mutat 21, 12–27.

Kong, Q.P., Yao, Y.G., Sun, C., Zhu, C.L., Zhong, L., Wang, C.Y., Cai, W.W., Xu, X.M., Xu, A.L., Zhang, Y.P., 2004. Phylogeographic analysis of mitochondrial DNA haplogroup F2 in China reveals T12338C in the initiation codon of the ND5 gene not to be pathogenic. J Hum Genet 49, 414–423.

Kovach, M., Sweeney, M., McCouch, S., 2007. New insights into the history of rice domestication. TRENDS in Genetics 23, 578–587.

Kraaijenbrink, T., van Driem, G.L., Opgenort, J.R.M.L., Tuladhar, N.M., de Knijff, P., 2007a. Allele frequency distribution for 21 autosomal STR loci in Nepal. Forensic Sci Int 168, 227–231.

Kraaijenbrink, T., van Driem, G.L., Tshering of Gaselô, K., de Knijff, P., 2007b. Allele frequency distribution for 21 autosomal STR loci in Bhutan. Forensic Sci Int 170, 68–72.

Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M., Paabo, S., 1997. Neandertal DNA sequences and the origin of modern humans. Cell 90, 19–30.

Krings, M., Geisert, H., Schmitz, R.W., Krainitzki, H., Paabo, S., 1999. DNA sequence of the mitochondrial hypervariable region II from the neandertal type specimen. Proc Natl Acad Sci U S A 96, 5581–555.

Krings, M., Capelli, C., Tschentscher, F., Geisert, H., Meyer, S., von Haeseler, A., Grossschmidt, K., Possnert, G., Paunovic, M., Pääbo, S., 2000. A view of Neandertal genetic diversity. Nat Genet 26, 144–146.

Kruglyak, L., 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22, 139–144.

Kulick, D., 1998. Anger, gender, language shift, and the politics of revelation in a Papua New Guinean village. Language ideologies: Practice and theory , 87–102.

Kumar, S., Nagarajan, M., Sandhu, J.S., Kumar, N., Behl, V., 2007. Phylogeography and domestication of Indian river buffalo. BMC Evol Biol 7, 186.

Kumar, S., Padmanabham, P.B.S.V., Ravuri, R.R., Uttaravalli, K., Koneru, P., Mukherjee, P.A., Das, B., Kotal, M., Xaviour, D., Saheb, S.Y., Rao, V.R., 2008. The earliest settlers' antiquity and evolutionary history of Indian populations: evidence from M2 mtDNA lineage. BMC Evol Biol 8, 230.

Kumar, S., Ravuri, R.R., Koneru, P., Urade, B.P., Sarkar, B.N., Chandrasekar, A., Rao, V.R., 2009. Reconstructing Indian-Australian phylogenetic link. BMC Evol Biol 9, 173.

Kumar, V., Reddy, M., 2003. Status of Austroasiatic groups in the peopling of India: An exploratory study based on the available prehistoric, linguistic and biological evidences. J. Biosci. 28, 507–522.

Kumar, V., Langstieh, B.T., Madhavi, K.V., Naidu, V.M., Singh, H.P., Biswas, S., Thangaraj, K., Singh, L., Reddy, B.M., 2006b. Global patterns in human mitochondrial DNA and Y-chromosome variation caused by spatial instability of the local cultural processes. PLoS Genet 2, e53.

Kumar, V., Langsiteh, B.T., Biswas, S., Babu, J.P., Rao, T.N., Thangaraj, K., Reddy, A.G., Singh, L., Reddy, B.M., 2006b. Asian and non-Asian origins of Mon-Khmer- and Mundari-speaking Austroasiatic populations of India. Am J Hum Biol 18, 461–469.

Kumar, V., Reddy, A.N.S., Babu, J.P., Rao, T.N., Langstieh, B.T., Thangaraj, K., Reddy, A.G., Singh, L., Reddy, B.M., 2007. Y-chromosome evidence suggests a common paternal heritage of Austroasiatic populations. BMC Evol Biol 7, 47.

Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L.G., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Silber, S., Oates, R., Rozen, S., Page, D.C., 2001. The AZFc region of the Y-chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nat Genet 29, 279–86.

Kutty, E., 2006. A glimpse on Vedic concepts Bharatiya Vidya Bhavan.

Lahr, M., Foley, R., 1994. Multiple dispersals and modern human origins. Evol Anthropol 3, 48–60.

Lahr, M.M., Foley, R.A., 1998. Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. Am J Phys Anthropol Suppl, 137–176.

Lal, B., 1997. The Earliest Civilization of South Asia: Rise, Maturity, and Decline Aryan Books International, New Delhi.

Lal, B., 2009. Deep Roots of Indian Civilization Aryan Book Internatonal, New Delhi.

Lalueza-Fox, C., Sampietro, M., Caramelli, D., Puder, Y., Lari, M., Calafell, F., Martinez-Maza, C., Bastir, M., Fortea, J., Rasilla, M., 2005. Neandertal evolutionary genetics: mitochondrial DNA data from the Iberian Peninsula. Mol Biol Evol 22, 1077.

Lao, O., Lu, T.T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., Balascakova, M., Bertranpetit, J., Bindoff, L.A., Comas, D., Holmlund, G., Kouvatsi, A., Macek, M., Mollet, I., Parson, W., Palo, J., Ploski, R., Sajantila, A., Tagliabracci, A., Gether, U., Werge, T., Rivadeneira, F., Hofman, A., Uitterlinden, A.G., Gieger, C., Wichmann, H.E., Rüther, A., Schreiber, S., Becker, C., Nürnberg, P., Nelson, M.R., Krawczak, M., Kayser, M., 2008. Correlation between genetic and geographic structure in Europe. Curr Biol 18, 1241–1248.

Larsen, C., 2002. Post-Pleistocene human evolution: bioarcheology of the agricultural transition, in: Ungar, P.S., Teaford, M.F. (Ed.), Human diet. Praeger Pub Text, pp. 206.

Lawler, A., 2008. Unmasking the Indus. Indus collapse: the end or the beginning of an Asian culture? Science 320, 1281–1283.

Lawler, A., 2010. A Forgotten Corridor Rediscovered. Science 328, 1092–1097.

Lewin, R., Foley, R., 2004. Principles of human evolution Blackwell Publishing Blackwell Science Ltd.

Lewis, M.P. (Ed.), 2009. Ethnologue: Languages of the World, Online version: http://www.ethnologue.com/, Sixteenth ed. SIL International, Dallas, Tex.

Li, H., Huang, Y., Mustavich, L.F., Zhang, F., Tan, J.Z., Wang, L.E., Qian, J., Gao, M.H., Jin, L., 2007. Y-chromosomes of prehistoric people along the Yangtze River. Hum Genet 122, 383–388.

Li, H., Wen, B., Chen, S.J., Su, B., Pramoonjago, P., Liu, Y., Pan, S., Qin, Z., Liu, W., Cheng, X., Yang, N., Li, X., Tran, D., Lu, D., Hsu, M.T., Deka, R., Marzuki, S., Tan, C.C., Jin, L., 2008. Paternal genetic affinity between Western Austronesians and Daic populations. BMC Evol Biol 8, 146.

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., Myers, R.M., 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319, 1100–1104.

Lieberman, E., Hauert, C., Nowak, M.A., 2005. Evolutionary dynamics on graphs. Nature 433, 312–316.

Loftus, R.T., MacHugh, D.E., Bradley, D.G., Sharp, P.M., Cunningham, P., 1994. Evidence for two independent domestications of cattle. Proc Natl Acad Sci U S A 91, 2757–2761.

Loogväli, E.L., Kivisild, T., Margus, T., Villems, R., 2009. Explaining the imperfection of the molecular clock of hominid mitochondria. PloS one 4, e8260.

Lutz, S., Weisser, H.J., Heizmann, J., Pollak, S., 1998. Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. Int J Legal Med 111, 67–77.

Lynch, M., 2010. Rate, molecular spectrum, and consequences of human mutation. Proc Natl Acad Sci U S A 107, 961–968.

Ma, J., Bennetzen, J., 2004. Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci U S A 101, 12404.

Macaulay, V., Richards, M., Sykes, B., 1999. Mitochondrial DNA recombination - no need to panic. Proc R Soc Lond B Biol Sci 266, 2037–2039.

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., Taha, A., Shaari, N.K., Raja, J.M., Ismail, P., Zainuddin, Z., Goodwin, W., Bulbeck, D., Bandelt, H.J., Oppenheimer, S., Torroni, A., Richards, M., 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science 308, 1034–1036.

MacDonald, G., 2009. Potential influence of the Pacific Ocean on the Indian summer monsoon and Harappan decline. Quaternary International, In Press, Corrected Proof, 1–9.

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

Majumder, P.P., 2001a. Ethnic populations of India as seen from an evolutionary perspective. J Biosci 26, 533–545.

Majumder, P.P., 2001b. Indian caste origins: genomic insights and future outlook. Genome Res 11, 931–92.

Mallory, J.P., 1989. In Search of the Indo-Europeans Thames and Hudson, London.

Malyarchuk, B.A., Grzybowski, T., Derenko, M.V., Czarny, J., Miscicka-Sliwka, D., 2006. Mitochondrial DNA diversity in the Polish Roma. Ann Hum Genet 70, 195–206.

Malyarchuk, B.A., Perkova, M.A., Derenko, M.V., Vanecek, T., Lazur, J., Gomolcak, P., 2008. Mitochondrial DNA variability in Slovaks, with application to the Roma origin. Ann Hum Genet 72, 228–240.

Mastana, S.S., Papiha, S.S., 1992. Origin of the Romany gypsies-genetic evidence. Z Morphol Anthropol 79, 43–51.
McAlpin, D., 1981. Proto-Elamo-Dravidian: the Evidence and its Implications American Philosophical Society, Philadelphia (PA).
McElreavey, K., Quintana-Murci, L., 2005. A population genetics perspective of the Indus Valley through uniparentally-inherited markers. Ann Hum Biol 32, 154–162.
McEvoy, B.P., Montgomery, G.W., McRae, A.F., Ripatti, S., Perola, M., Spector, T.D., Cherkas, L., Ahmadi, K.R., Boomsma, D., Willemsen, G., Hottenga, J.J., Pedersen, N.L., Magnusson, P.K.E., Kyvik, K.O., Christensen, K., Kaprio, J., Heikkilä, K., Palotie, A., Widen, E., Muilu, J., Syvänen, A.C., Liljedahl, U., Hardiman, O., Cronin, S., Peltonen, L., Martin, N.G., Visscher, P.M., 2009. Geographical structure and differential natural selection among North European populations. Genome Res 19, 804–814.
Meadow, R., 2004. Notes on the faunal remains from Mehrgarh, with focus on cattle (Bos), in: Allchin, B. (Ed.), South Asian archaeology. Cambridge University Press, , pp. 30–40.
Mellars, P., 2004. Neanderthals and the modern human colonization of Europe. Nature 432, 461–465.
Mellars, P., 2006a. A new radiocarbon revolution and the dispersal of modern humans in Eurasia. Nature 439, 931–935.
Mellars, P., 2006b. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. Science 313, 796–800.
Merriwether, D.A., Clark, A.G., Ballinger, S.W., Schurr, T.G., Soodyall, H., Jenkins, T., Sherry, S.T., Wallace, D.C., 1991. The structure of human mitochondrial DNA variation. J Mol Evol 33, 543–555.
Merriwether, D.A., Hodgson, J.A., Friedlaender, F.R., Allaby, R., Cerchio, S., Koki, G., Friedlaender, J.S., 2005. Ancient mitochondrial M haplogroups identified in the Southwest Pacific. Proc Natl Acad Sci U S A 102, 13034–13039.
Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T., Endicott, P., Mastana, S., Papiha, S.S., Skorecki, K., Torroni, A., Villems, R., 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet 5, 26.
Metspalu, M., 2005. Through the course of prehistory in India: tracing the mtDNA trail, PhD, University of Tartu, Estonia.
Metspalu, M., Kivisild, T., Bandelt, H.J., Richards, M., Villems, R., 2006. The pioneer settlement of modern humans in Asia, in: Bandelt, H.J., Macaulay, V., Richards, M. (Ed.), Human mitochondrial DNA and the evolution of Homo sapiens. Springer-Verlag, Heidelberg
Mihaylova, V., Hantke, J., Sinigerska, I., Cherninkova, S., Raicheva, M., Bouwer, S., Tincheva, R., Khuyomdziev, D., Bertranpetit, J., Chandler, D., Angelicheva, D., Kremensky, I., Seeman, P., Tournev, I., Kalaydjieva, L., 2007. Highly variable neural involvement in sphingomyelinase-deficient Niemann-Pick disease caused by an ancestral Gypsy mutation. Brain 130, 1050–1061.
Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., Sukernik, R.I., Olckers, A., Wallace, D.C., 2003. Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA 100, 171–176.
Mishra, V., 2001. human colonization of India. J Biosci 26 supp.

Mohyuddin, A., Ayub, Q., Underhill, P.A., Tyler-Smith, C., Mehdi, S.Q., 2006. Detection of novel Y SNPs provides further insights into Y chromosomal variation in Pakistan. J Hum Genet 51, 375–378.

Mountain, J.L., Hebert, J.M., Bhattacharyya, S., Underhill, P.A., Ottolenghi, C., Gadgil, M., Cavalli-Sforza, L.L., 1995. Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56, 979–992.

Mundlay, A., 1996. "Who are the Nihals? What Do They Speak.". Mother Tongue 2, 5–9.

Nachman, M.W., Crowell, S.L., 2000. Estimate of the mutation rate per nucleotide in humans. Genetics 156, 297–304.

Nass, S., Nass, M.M.K., 1963. INTRAMITOCHONDRIAL FIBERS WITH DNA CHARACTERISTICS: II. Enzymatic and Other Hydrolytic Treatments 10.1083/jcb.19.3.613. J. Cell Biol. 19, 613–629.

Naugler, C., 2008. Hemochromatosis: a Neolithic adaptation to cereal grain diets. Med Hypotheses 70, 691–692.

Nei, M., Roychoudhury, A., 1982. Genetic relationship and evolution of human races. Evol Biol 14, 1–59.

Neiman, M., Taylor, D.R., 2009. The causes of mutation accumulation in mitochondrial genomes. Proc Biol Sci 276, 1201–1209.

Nelis, M., Esko, T., Mägi, R., Zimprich, F., Zimprich, A., Toncheva, D., Karachanak, S., Piskácková, T., Balascák, I., Peltonen, L., Jakkula, E., Rehnström, K., Lathrop, M., Heath, S., Galan, P., Schreiber, S., Meitinger, T., Pfeufer, A., Wichmann, H.E., Melegh, B., Polgár, N., Toniolo, D., Gasparini, P., D'Adamo, P., Klovins, J., Nikitina-Zake, L., Kucinskas, V., Kasnauskiene, J., Lubinski, J., Debniak, T., Limborska, S., Khrunin, A., Estivill, X., Rabionet, R., Marsal, S., Julià, A., Antonarakis, S.E., Deutsch, S., Borel, C., Attar, H., Gagnebin, M., Macek, M., Krawczak, M., Remm, M., Metspalu, A., 2009. Genetic structure of Europeans: a view from the North-East. PloS one 4, e5472.

Nishank, S.S., Chhotray, G.P., Kar, S.K., Ranjit, M.R., 2008. Molecular variants of G6PD deficiency among certain tribal communities of Orissa, India. Ann Hum Biol 35, 355–361.

Noonan, J.P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Pääbo, S., Pritchard, J.K., Rubin, E.M., 2006. Sequencing and analysis of Neanderthal genomic DNA. Science 314, 1113–1118.

Nordborg, M., 1998. On the probability of Neanderthal ancestry. The American Journal of Human Genetics 63, 1237–1240.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R., Stephens, M., Bustamante, C.D., 2008. Genes mirror geography within Europe. Nature 456, 98–101.

Novembre, J., Pritchard, J.K., Coop, G., 2007. Adaptive drool in the gene pool. Nat Genet 39, 1188–1190.

Núñez, L., Grosjean, M., Cartajena, I., 2002. Human occupations and climate change in the Puna de Atacama, Chile. Science 298, 821–824.

O'Connell, J., Allen, J., 2007. Pre-LGM Sahul (Pleistocene Australia-New Guinea) and the Archaeology of Early Modern Humans, in: Mellars, P., Boyle, K., Bar-Yosef, O. (Ed.), Rethinking the human revolution. McDonald Inst of Archeological, pp. 436.

Ohno, S., 1967. Sex chromosomes and sex-linked genes Springer, Berlin.

Oleksyk, T.K., Smith, M.W., O'Brien, S.J., 2010. Genome-wide scans for footprints of natural selection. Philos Trans R Soc Lond B Biol Sci 365, 185–205.

Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Al-Zahery, N., Scozzari, R., Cruciani, F., Behar, D.M., Dugoujon, J.M., Coudray, C., Santachiara-Benerecetti, A.S., Semino, O., Bandelt, H.J., Torroni, A., 2006. The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science 314, 1767–1770.

Olivo, P.D., Van de Walle, M.J., Laipis, P.J., Hauswirth, W.W., 1983. Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. Nature 306, 400–402.

Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T., Stoneking, M., 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet 29, 20–21.

Ottenheimer, M., 1996. Forbidden relatives: The American myth of cousin marriage University of Illinois Press

Ovchinnikov, I.V., Gotherstrom, A., Romanova, G.P., Kharitonov, V.M., Liden, K., Goodwin, W., 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature 404, 490–43.

Overpeck, J., Anderson, D., Trumbore, S., Prell, W., 1996. The southwest Indian monsoon over the last 18000 years. Climate Dyn. 12, 213–225.

Palanichamy, M., Sun, C., Agrawal, S., Bandelt, H.J., Kong, Q.P., Khan, F., Wang, C.Y., Chaudhuri, T., Palla, V., Zhang, Y.P., 2004. Phylogeny of mtDNA macrohaplogroup N in India based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75, 966–978.

Palanichamy, M.G., Agrawal, S., Yao, Y.G., Kong, Q.P., Sun, C., Khan, F., Chaudhuri, T.K., Zhang, Y.P., 2006. Comment on "Reconstructing the origin of Andaman islanders". Science 311, 470; author reply 470.

Papiha, S.S., 1996. Genetic variation in India. Hum Biol 68, 607–628.

Parkin, E.J., Kraayenbrink, T., van Driem, G.L., Tshering Of Gaselô, K., de Knijff, P., Jobling, M.A., 2006. 26-Locus Y-STR typing in a Bhutanese population sample. Forensic Sci Int 161, 1–7.

Parkin, E.J., Kraayenbrink, T., Opgenort, J.R.M.L., van Driem, G.L., Tuladhar, N.M., de Knijff, P., Jobling, M.A., 2007. Diversity of 26-locus Y-STR haplotypes in a Nepalese population sample: isolation and drift in the Himalayas. Forensic Sci Int 166, 176–181.

Parpola, A. (Ed.), 1988. The coming of the Aryans to Iran and India and the cultural and ethnic identity of the Dāsas.

Passarino, G., Semino, O., Bernini, L.F., Santachiara-Benerecetti, A.S., 1996. Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. Am J Hum Genet 59, 927–934.

Patnaik, R., Chauhan, P., 2009. India at the cross-roads of human evolution. J Biosci 34, 729.

Patnaik, R., Chauhan, P.R., Rao, M., Blackwell, B., Skinner, A., Sahni, A., Chauhan, M., Khan, H., 2009. New geochronological, paleoclimatological, and archaeological data from the Narmada Valley hominin locality, central India. J Hum Evol 56, 114–133.

Peng, Y., Shi, H., Qi, X.B., Xiao, C.J., Zhong, H., Ma, R.L.Z., Su, B., 2010. The ADH1B Arg47His polymorphism in east Asian populations and expansion of rice domestication in history. BMC Evol Biol 10, 15.

Perry, G.H., Dominy, N.J., Claw, K.G., Lee, A.S., Fiegler, H., Redon, R., Werner, J., Villanea, F.A., Mountain, J.L., Misra, R., Carter, N.P., Lee, C., Stone, A.C., 2007. Diet and the evolution of human amylase gene copy number variation. Nat Genet 39, 1256–1260.

Petraglia MD, Allchin B. 2007. In The evolution and history of human populations in South Asia, ed. MD Petraglia, B Allchin pp. 464. : Springer Verlag.

Petraglia, M., Clarkson, C., Boivin, N., Haslam, M., Korisettar, R., Chaubey, G., Ditchfield, P., Fuller, D., James, H., Jones, S., Kivisild, T., Koshy, J., Lahr, M.M., Metspalu, M., Roberts, R., Arnold, L., 2009. Population increase and environmental deterioration correspond with microlithic innovations in South Asia ca. 35,000 YBP. Proc Natl Acad Sci U S A 106, 12261–12266.

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

Poongothai, J., Gopenath, T.S., Manonayaki, S., 2009. Genetics of human male infertility. Singapore Med J 50, 336–347.

Possehl, G., 1979. Ancient cities of the Indus. Carolina Academic Press.

Powell, A., Shennan, S., Thomas, M.G., 2009. Late Pleistocene demography and the appearance of modern human behavior. Science 324, 1298–1301.

Prasad, B.V., Ricker, C.E., Watkins, W.S., Dixon, M.E., Rao, B.B., Naidu, J.M., Jorde, L.B., Bamshad, M., 2001. Mitochondrial DNA variation in Nicobarese Islanders. Hum Biol 73, 715–725.

Prashanth, S., Parani, M., Mohanty, B., Talame, V., Tuberosa, R., Parida, A., 2002. Genetic diversity in cultivars and landraces of Oryza sativa subsp. indica as revealed by AFLP markers. Genome 45, 451–459.

Pratap, V., 2001. Vasudhaiva Kutumbakam: A New Alliance for Democracy in the Era of Globalisation. Lokayan Bulletin.

Prufer, K., Stenzel, U., Hofreiter, M., Paabo, S., Kelso, J., Green, R., 2010. Computational challenges in the analysis of ancient DNA. Genome Biol 11, R47.

Puhvel, J. (Ed.), 2004. Hittite Etymological Dictionary Words beginning with M Mouton De Gruyter.

Purana, V., 1961. Gita Press, Gorakhpur. English translation by HH Wilson, Calcutta.

Puri, V., 2008. Vedic Sarasvati: Scientific Signatures on its Origin from the Himalaya, in: Kalyanaraman, S., Sarasvati Research and Education Trust (Madras, India) (Ed.), Vedic River Sarasvati and Hindu civilization. New Delhi, pp. 398.

Qamar, R., Ayub, Q., Khaliq, S., Mansoor, A., Karafet, T., Mehdi, S.Q., Hammer, M.F., 1999. African and Levantine origins of Pakistani YAP+ Y-chromosomes. Hum Biol 71, 745–755.

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

Quintana-Murci, L., Semino, O., Bandelt, H.J., Passarino, G., McElreavey, K., Santachiara-Benerecetti, A.S., 1999. Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat Genet 23, 437–441.

Quintana-Murci, L., Krausz, C., Zerjal, T., Sayar, S.H., Hammer, M.F., Mehdi, S.Q., Ayub, Q., Qamar, R., Mohyuddin, A., Radhakrishna, U., Jobling, M.A., Tyler-Smith, C., McElreavey, K., 2001. Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. Am J Hum Genet 68, 537–542.

Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A.S., Coppa, A., Ayub, Q., Mohyuddin, A., Tyler-Smith, C., Qasim Mehdi, S., Torroni, A., McElreavey, K., 2004. Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. Am J Hum Genet 74, 827–845.

Radcliffe-Brown, A.R. (Ed.), 1948. The Andaman Islanders.
Radhakrishna BP, 1999. Vedic Sarasvati and the dawn of Indian civilization, in: Radhakrishna, B.P., Merh, S.S. (Ed.), Vedic Sarasvati. pp. 329.
Radhakrishna, B., 1999. Vedic Sarasvati: Evolutionary History of a Lost River of Northwestern India, in: Radhakrishna, B.P., Merh, S.S. (Ed.), Vedic Sarasvati. Mem. Geol. Soc. of India, , pp. 5–13.
Radhakrishnan, R., 1981. The Nancowry word: Phonology, affixal morphology, and roots of a Nicobarese language Linguistic Research.
Ramana, G.V., Su, B., Jin, L., Singh, L., Wang, N., Underhill, P., Chakraborty, R., 2001. Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. Eur J Hum Genet 9, 695–700.
Rao, R.P.N., Yadav, N., Vahia, M.N., Joglekar, H., Adhikari, R., Mahadevan, I., 2009. Entropic evidence for linguistic structure in the Indus script. Science 324, 1165.
Redd, A.J., Takezaki, N., Sherry, S.T., McGarvey, S.T., Sofro, A.S., Stoneking, M., 1995. Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12, 604–615.
Reddy, B.M., Naidu, V.M., Madhavi, V.K., Thangaraj, L.K., Kumar, V., Langstieh, B.T., Venkatramana, P., Reddy, A.G., Singh, L., 2005. Microsatellite diversity in Andhra Pradesh, India: genetic stratification versus social stratification. Hum Biol 77, 803–823.
Reddy, B.M., Langstieh, B.T., Kumar, V., Nagaraja, T., Reddy, A.N., Meka, A., Reddy, A.G., Thangaraj, K., Singh, L., 2007. Austroasiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. PLoS ONE 2, e1141.
Reich, D., Thangaraj, K., Patterson, N., Price, A.L., Singh, L., 2009. Reconstructing Indian population history. Nature 461, 489–494.
Reijo, R., Lee, T.Y., Salo, P., Alagappan, R., Brown, L.G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O., 1995. Diverse spermatogenic defects in humans caused by Y-chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet 10, 383–393.
Reijo, R.A., Dorfman, D.M., Slee, R., Renshaw, A.A., Loughlin, K.R., Cooke, H., Page, D.C., 2000. DAZ family proteins exist throughout male germ cell development and transit from nucleus to cytoplasm at meiosis in humans and mice. Biol Reprod 63, 1490–1496.
Renfrew, A., 1996. Language families and the spread of farming, in: Harris, C. (Ed.), The Origins and Spread of Agriculture and Pastoralism. UCL Press, London
Renfrew, C., 1991. The coming of the Aryans to Iran and India and the cultural and ethnic identity of the D?sas. By Asko Parpola. (Studia Orientalia, Vol. 64.) pp. 195?302, 33 figs. Helsinki, The Finnish Oriental Society, 1988. Journal of the Royal Asiatic Society (Third Series) 1, 106–109.
Renfrew, C., 1996. Languages families and the spread of farming, in: Harris, D.R. (Ed.), The origins and spread of agriculture and pastoralism in Eurasia. Routledge, , pp. 594.
Renfrew, C., 2000. At the edge of knowability: towards a prehistory of languages. Cambridge Archaeological Journal 10, 7–34.
Renfrew, C., Boyle, K., 2000. Archaeogenetics: DNA and the population prehistory of Europe. (McDonald Institute Monographs) McDonald Institute for Archaeological Research., Cambridge.

Repping, S., Skaletsky, H., Lange, J., Silber, S., Van Der Veen, F., Oates, R.D., Page, D.C., Rozen, S., 2002. Recombination between palindromes P5 and P1 on the human Y-chromosome causes massive deletions and spermatogenic failure. Am J Hum Genet 71, 906–22.

Repping, S., Skaletsky, H., Brown, L., van Daalen, S.K., Korver, C.M., Pyntikova, T., Kuroda-Kawaguchi, T., de Vries, J.W., Oates, R.D., Silber, S., van der Veen, F., Page, D.C., Rozen, S., 2003. Polymorphism for a 1.6-Mb deletion of the human Y-chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet 35, 247–51. Epub 2003 Oct 5.

Repping, S., van Daalen, S.K.M., Brown, L.G., Korver, C.M., Lange, J., Marszalek, J.D., Pyntikova, T., van der Veen, F., Skaletsky, H., Page, D.C., Rozen, S., 2006. High mutation rates have driven extensive structural polymorphism among human Y-chromosomes. Nat Genet 38, 463–467.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Gölge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Calì, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Nørby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozari, R., Torroni, A., Bandelt, H.J., 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67, 1251–1276.

Richards, M.P., Schulting, R.J., Hedges, R.E.M., 2003. Archaeology: sharp shift in diet at onset of Neolithic. Nature 425, 366.

Richards, M., Bandelt, H.J., Kivisild, T., Oppenheimer, S., 2006. A model for the dispersal of modern humans out of Africa, in: Bandelt, H.J., Macaulay, V., Richards, M. (Ed.), Human mitochondrial DNA and the evolution of Homo sapiens. Springer-Verlag, Heidelberg.

Rightmire, G., 2001. Comparison of Middle Pleistocene hominids from Africa and Asia, in: Barham, L., Robson-Brown, L., Robson-Brown, K. (Ed.), Human Roots: Africa and Asia in the Middle Pleistocene. Western Academic Press, Bristol, pp. 123–135.

Rispe, C., Moran, N., 2000. Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. American Naturalist 156, 1431–1440.

Rochow, I., Matschke, K., 1998. Neues zu den zigeunern im Byzantinischen reich um die wende von 13. Zum 14. Jahrhundert. Jahrbuch der O¨ sterreichischen Byzantinistik , 241–254.

Rohde, D.L.T., Olson, S., Chang, J.T., 2004. Modelling the recent common ancestry of all living humans. Nature 431, 562–566.

Rootsi, S., 2004. Human Y-chromosomal variation in European populations, PhD, Human Y-chromosomal variation in European populations, Estonia.

Rootsi, S., Zhivotovsky, L.A., Baldovic, M., Kayser, M., Kutuev, I.A., Khusainova, R., Bermisheva, M.A., Gubina, M., Fedorova, S.A., Ilumäe, A.M., Khusnutdinova, E.K., Voevoda, M.I., Osipova, L.P., Stoneking, M., Lin, A.A., Ferak, V., Parik, J., Kivisild, T., Underhill, P.A., Villems, R., 2007. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. Eur J Hum Genet 15, 204–211.

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

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., Feldman, M.W., 2002. Genetic structure of human populations. Science 298, 2381–2385.

Rosser, Z.H., Zerjal, T., Hurles, M.E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D.G., Brede, G., Cooper, G., Corte-Real, H.B., de Knijff, P., Decorte, R., Dubrova, Y.E., Evgrafov, O., Gilissen, A., Glisic, S., Golge, M., Hill, E.W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S.A., Krumina, A., Kucinskas, V., Lavinha, J., Livshits, L.A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T.A., Mikelsaar, A.V., Mitchell, R.J., Nafa, K., Nicholson, J., Norby, S., Pandya, A., Parik, J., Patsalis, P.C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M.J., Previdere, C., Roewer, L., Rootsi, S., Rubinsztein, D.C., Saillard, J., Santos, F.R., Stefanescu, G., Sykes, B.C., Tolun, A., Villems, R., Tyler-Smith, C., Jobling, M.A., 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67, 1526–1543.

Roychoudhury, S., Roy, S., Basu, A., Banerjee, R., Vishwanathan, H., Usha Rani, M.V., Sil, S.K., Mitra, M., Majumder, P.P., 2001. Genomic structures and population histories of linguistically distinct tribal groups of India. Hum Genet 109, 339–350.

Roze, D., Rousset, F., Michalakis, Y., 2005. Germline bottlenecks, biparental inheritance and selection on mitochondrial variants: a two-level selection model. Genetics 170, 1385–1399.

Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Page, D.C., 2003. Abundant gene conversion between arms of palindromes in human and ape Y-chromosomes. Nature 423, 873–86.

Rozen, S., Marszalek, J.D., Alagappan, R.K., Skaletsky, H., Page, D.C., 2009. Remarkably little variation in proteins encoded by the Y-chromosome's single-copy genes, implying effective purifying selection. Am J Hum Genet 85, 923–928.

Ruggiu, M., Speed, R., Taggart, M., McKay, S.J., Kilanowski, F., Saunders, P., Dorin, J., Cooke, H.J., 1997. The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis. Nature 389, 73–77.

Ruhlen, M., 1991. A guide to the world's languages Edward Arnold, London.

Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., Wallace, D.C., 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. Science 303, 223–26.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S., 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832–837.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S., Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Sun, W., Wang, H., Wang, Y., Xiong,

X., Xu, L., Waye, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913–918.

Sahoo, S., Singh, A., Himabindu, G., Banerjee, J., Sitalaximi, T., Gaikwad, S., Trivedi, R., Endicott, P., Kivisild, T., Metspalu, M., Villems, R., Kashyap, V.K., 2006. A prehistory of Indian Y-chromosomes: evaluating demic diffusion scenarios. Proc Natl Acad Sci U S A 103, 843–848.

Saillard, J., Forster, P., Lynnerup, N., Bandelt, H.J., Norby, S., 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 67, 718–726.

Salmela, E., Lappalainen, T., Fransson, I., Andersen, P.M., Dahlman-Wright, K., Fiebig, A., Sistonen, P., Savontaus, M.L., Schreiber, S., Kere, J., Lahermo, P., 2008. Genome-wide analysis of single nucleotide polymorphisms uncovers population structure in Northern Europe. PloS one 3, e3519.

Saraswat, K., 1992. Archaeobotanical remains in ancient cultural and socio-economical dynamics of the Indian subcontinent. Palaeobotanist 40, 514–545.

Saraswat, K., 1993. Plant economy of late Harappan at Hulas. Puratattva 23, 1–12.

Saraswat, K., 2004. Plant economy of early farming communities, in: Singh, B. (Ed.), Early Farming Communities of the Kaimur (Excavations at Senuwar). Jaipur: Publication Scheme, , pp. 416–535.

Saraswati, C., Bhavan, B., 1995. Hindu Dharma: The Universal Way of Life Bharatiya Vidya Bhavan

Sareen, J., 2005. The Lost Tribes of India, http://www.swaveda.com/articles.php?action=show&id=31

Saxena, R., Brown, L.G., Hawkins, T., Alagappan, R.K., Skaletsky, H., Reeve, M.P., Reijo, R., Rozen, S., Dinulos, M.B., Disteche, C.M., Page, D.C., 1996. The DAZ gene cluster on the human Y-chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. Nat Genet 14, 292–299.

Saxena, R., de Vries, J.W., Repping, S., Alagappan, R.K., Skaletsky, H., Brown, L.G., Ma, P., Chen, E., Hoovers, J.M., Page, D.C., 2000. Four DAZ genes in two clusters found in the AZFc region of the human Y-chromosome. Genomics 67, 256–267.

Seielstad, M.T., Minch, E., Cavalli-Sforza, L.L., 1998. Genetic evidence for a higher female migration rate in humans. Nat Genet 20, 278–280.

Semino, O., Torroni, A., Scozzari, R., Brega, A., Santachiara Benerecetti, A.S., 1991. Mitochondrial DNA polymorphisms among Hindus: a comparison with the Tharus of Nepal. Ann Hum Genet 55, 123–136.

Sengupta, S., Zhivotovsky, L.A., King, R., Mehdi, S.Q., Edmonds, C.A., Chow, C.E., Lin, A.A., Mitra, M., Sil, S.K., Ramesh, A., Usha Rani, M.V., Thakur, C.M., Cavalli-Sforza, L.L., Majumder, P.P., Underhill, P.A., 2006. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. Am J Hum Genet 78, 202–221.

Serre, D., Langaney, A., Chech, M., Teschler-Nicola, M., Paunovic, M., Mennecier, P., Hofreiter, M., Possnert, G.G., Paabo, S., 2004. No Evidence of Neandertal mtDNA Contribution to Early Modern Humans. PLoS Biol 2, E57. Epub 2004 Mar 16.

Shadan, S., 2007. Evolutionary genetics: you are what you ate. Nature 449, 155.

Sharma, S., Joachimski, M., Sharma, M., Tobschall, H., Singh, I., Sharma, C., Chauhan, M., Morgenroth, G., 2004. Lateglacial and Holocene environmental changes in Ganga plain, Northern India. BMC Biol 23, 145–159.

Sharma, S., Saha, A., Rai, E., Bhat, A., Bamezai, R., 2005. Human mtDNA hypervariable regions, HVR I and II, hint at deep common maternal founder and subsequent maternal gene flow in Indian population groups. J Hum Genet 50, 497–506.

Shinde, V., Deshpande, S., Osada, T., Uno, T., 2009. Basic Issues in Harappan Archaeology: Some Thoughts. Ancient Asia 1.

Singh, G., 1971. The Indus Valley culture (seen in the context of postglacial climate and ecological studies in Northwest India). Archaeol. Phys. Anthropol. Oceania 6, 177–189.

Singh, G., Joshi, R., Singh, A., 1973. Pollen-rain from the vegetation of northwest India. New Phytologist 72, 191–206.

Singh, K.S., 1997. People of India Oxford University Press, Oxford.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S.F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S.P., Waterston, R.H., Wilson, R.K., Rozen, S., Page, D.C., 2003. The male-specific region of the human Y-chromosome is a mosaic of discrete sequence classes. Nature 423, 825–37.

Slee, R., Grimes, B., Speed, R.M., Taggart, M., Maguire, S.M., Ross, A., McGill, N.I., Saunders, P.T., Cooke, H.J., 1999. A human DAZ transgene confers partial rescue of the mouse Dazl null phenotype. Proc Natl Acad Sci U S A 96, 8040–8045.

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A., Salas, A., Oppenheimer, S., Macaulay, V., Richards, M.B., 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. Am J Hum Genet 84, 740–759.

Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H.J., Torroni, A., Richards, M.B., 2010. The Archaeogenetics of Europe. Curr Biol 20, R174-R183.

Soejima, M., Tachida, H., Ishida, T., Sano, A., Koda, Y., 2006. Evidence for Recent Positive Selection at the Human AIM1 Locus in a European Population 10.1093/molbev/msj018. Mol Biol Evol 23, 179–188.

Soltis, P.S., Soltis, D.E., Savolainen, V., Crane, P.R., Barraclough, T.G., 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. Proc Natl Acad Sci U S A 99, 4430–4435.

Sonakia, A., 1984. The scull-cap of Early Man and Associated Mammalin Fauna from Narmada Valley Alluvium, hoshangabad Area, Madhya Pradesh (India). Records of the Geological Survey of India , 159–172.

Soodyall, H., Jenkins, T., 1992. Mitochondrial DNA polymorphisms in Khoisan populations from southern Africa. Ann Hum Genet 56, 315–324.

Sreenath, J., Ahmad, S., 1989. All India Anthropometric Survey, South Zone: Analysis of Data. Anthropological Survey of India.

Staubwasser, M., Sirocko, F., Grootes, P., Segl, M., 2003. Climate change at the 4.2 ka BP termination of the Indus Valley Civilization and Holocene South Asian monsoon variability. Geophys. Res. Lett. 30, 71–74.

Steinemann, M., Steinemann, S., 1998. Enigma of Y-chromosome degeneration: neo-Y and neo-X chromosomes of Drosophila miranda a model for sex chromosome evolution. Genetica 102–103, 409–420.
Stewart, J.B., Freyer, C., Elson, J.L., Wredenberg, A., Cansu, Z., Trifunovic, A., Larsson, N.G., 2008. Strong purifying selection in transmission of mammalian mitochondrial DNA. PLoS Biol 6, e10.
Stoneking, M., Soodyall, H., 1996. Human evolution and the mitochondrial genome. Curr Opin Genet Dev 6, 731–736.
Stringer, C., 2000. Coasting out of Africa. Nature 405, 24–5, 27.
Stringer, C.B., Andrews, P., 1988. Genetic and fossil evidence for the origin of modern humans. Science 239, 1263–1268.
Sun, C., Skaletsky, H., Rozen, S., Gromoll, J., Nieschlag, E., Oates, R., Page, D.C., 2000. Deletion of azoospermia factor a (AZFa) region of human Y-chromosome caused by recombination between HERV15 proviruses. Hum Mol Genet 9, 2291–226.
Sun, C., Kong, Q.P., Palanichamy, M.G., Agrawal, S., Bandelt, H.J., Yao, Y.G., Khan, F., Zhu, C.L., Chaudhuri, T.K., Zhang, Y.P., 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. Mol Biol Evol 23, 683–690.
Sun, C., Kong, Q.P., Zhang, Y.P., 2007. The role of climate in human mitochondrial DNA evolution: a reappraisal. Genomics 89, 338–342.
Talageri, S., 2000. The Rigveda: a historical analysis Aditya Prakashan, India.
Tambets, K., 2004. Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: A phylogeographic approach, PhD, Tartu University, Estonia.
Tan, L., Li, X., Liu, F., Sun, X., Li, C., Zhu, Z., Fu, Y., Cai, H., Wang, X., Xie, D., Sun, C., 2008. Control of a key transition from prostrate to erect growth in rice domestication. Nat Genet 40, 1360–1364.
Templeton, A., 2002. Out of Africa again and again. Nature 416, 45–51.
Templeton, A., 2005. Haplotype trees and modern human origins. Yearbook of Physical Anthropology 48, 33.
Terreros, M.C., Rowold, D., Luis, J.R., Khan, F., Agrawal, S., Herrera, R.J., 2007. North Indian Muslims: enclaves of foreign DNA or Hindu converts? Am J Phys Anthropol 133, 1004–1012.
Thangaraj, K., Ramana, G.V., Singh, L., 1999. Y-chromosome and mitochondrial DNA polymorphisms in Indian populations. Electrophoresis 20, 1743–1747.
Thangaraj, K., Singh, L., Reddy, A.G., Rao, V.R., Sehgal, S.C., Underhill, P.A., Pierson, M., Frame, I.G., Hagelberg, E., 2003a. Genetic affinities of the Andaman Islanders, a vanishing human population. Curr Biol 13, 86–93.
Thangaraj, K., Gupta, N.J., Pavani, K., Reddy, A.G., Subramainan, S., Rani, D.S., Ghosh, B., Chakravarty, B., Singh, L., 2003b. Y-chromosome deletions in azoospermic men in India. J Androl 24, 588–597.
Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A.G., Singh, V.K., Rasalkar, A.A., Singh, L., 2005a. Reconstructing the origin of Andaman Islanders. Science 308, 996.
Thangaraj, K., Sridhar, V., Kivisild, T., Reddy, A.G., Chaubey, G., Singh, V.K., Kaur, S., Agarawal, P., Rai, A., Gupta, J., Mallick, C.B., Kumar, N., Velavan, T.P., Suganthan, R., Udaykumar, D., Kumar, R., Mishra, R., Khan, A., Annapurna, C., Singh, L., 2005b. Different population histories of the Mundari- and Mon-Khmer-speaking Austroasiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. Hum Genet 116, 507–517.

Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A.G., Singh, V.K., Rasalkar, A.A., Singh, L., 2006a. Response to comment on "Reconstructing the origin of Andaman islanders". Science 311.

Thangaraj, K., Chaubey, G., Reddy, A.G., Singh, V.K., Singh, L., 2006b. Unique origin of Andaman Islanders: insight from autosomal loci. J Hum Genet 51, 800–804.

Thangaraj, K., Chaubey, G., Reddy, A.G., Singh, V.K., Singh, L., 2007. Autosomal STR data on the enigmatic Andaman Islanders. Forensic Sci Int 169, 247–251.

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

Thangaraj, K., Nandan, A., Sharma, V., Sharma, V.K., Eaaswarkhanth, M., Patra, P.K., Singh, S., Rekha, S., Dua, M., Verma, N., Reddy, A.G., Singh, L., 2009. Deep rooting in-situ expansion of mtDNA Haplogroup R8 in South Asia. PloS one 4, e6545.

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

Thapar, R.,1966. A History of India, Vol. 1.Penguin.

Thomas, P., Joglekar, P., 1994. Holocene faunal studies. Man and Environment 19, 179–203.

Tishkoff, S.A., Dietzsch, E., Speed, W., Pakstis, A.J., Kidd, J.R., Cheung, K., Bonne-Tamir, B., Santachiara-Benerecetti, A.S., Moral, P., Krings, M., 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271, 1380–1387.

Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M., Ibrahim, M., Omar, S.A., Lema, G., Nyambo, T.B., Ghori, J., Bumpstead, S., Pritchard, J.K., Wray, G.A., Deloukas, P., 2007. Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 39, 31–40.

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

Tripathi, J., Tripathi, K., Bock, B., Rajamani, V., Eisenhauer, A., 2004. "Is River Ghaggar, Saraswati? Geochemical Constraints,". Curr Sci 87.

Trivedi, R., Sitalaximi, T., Banerjee, J., Singh, A., Sircar, P.K., Kashyap, V.K., 2006. Molecular insights into the origins of the Shompen, a declining population of the Nicobar archipelago. J Hum Genet 51, 217–226.

Trivedi, R., Sahoo, S., Singh, A., Bindu, G., Banerjee, J., Tandon, M., Gaikwad, S., Rajkumar, R., Sitalaximi, T., Ashma, R., 2008. Genetic Imprints of Pleistocene Origin of Indian Populations: A Comprehensive Phylogeographic Sketch of Indian Y-Chromosomes. Int J Hum Genet 8, 97–118.

Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykes, B.C., Bradley, D.G., 2001. Genetic evidence for Near-Eastern origins of European cattle. Nature 410, 1088–1091.

Umetsu, K., Yuasa, I., Yamashita, T., Saito, S., Yamaguchi, T., Ellepola, S.B., Ishida, T., Suzuki, T., 1989. Genetic polymorphisms of orosomucoid and alpha-2-HS-glycoprotein in Thai, Sri Lankan and Paraguayan populations. Jinrui Idengaku Zasshi 34, 195–202.

Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W., Kauffman, E., Bonne-tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J., Mehdi, S., Seielstad, M., Wells, R., Piazza, A., Davis, R., Feldman, M., Cavalli-Sforza, L.L., Oefner, P.J., 2000. Y-chromosome sequence variation and the history of human populations. Nat Genet 26, 358–361.

Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Mirazon Lahr, M., Foley, R., Oefner, P.J., Cavalli-Sforza, L.L., 2001. The phylogeography of Y-chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet. 65, 43–62.

Underhill, P.A., Kivisild, T., 2007. Use of Y-chromosome and mitochondrial DNA population structure in tracing human migrations. Annu Rev Genet 41, 539–564.

Underhill, P.A., Myres, N.M., Rootsi, S., Metspalu, M., Zhivotovsky, L.A., King, R.J., Lin, A.A., Chow, C.E.T., Semino, O., Battaglia, V., Kutuev, I., Järve, M., Chaubey, G., Ayub, Q., Mohyuddin, A., Mehdi, S.Q., Sengupta, S., Rogaev, E.I., Khusnutdinova, E.K., Pshenichnov, A., Balanovsky, O., Balanovska, E., Jeran, N., Augustin, D.H., Baldovic, M., Herrera, R.J., Thangaraj, K., Singh, V., Singh, L., Majumder, P., Rudan, P., Primorac, D., Villems, R., Kivisild, T., 2010. Separating the post-Glacial coancestry of European and Asian Y-chromosomes within haplogroup R1a. Eur J Hum Genet 18,479–84.

Valdiya, K., 1996. River piracy. Resonance 1, 19–28.

Valdiya, K., 2008. River Piracy: Sarasvati that Disappeared, in: Kalyanaraman, S., Sarasvati Research and Education Trust (Madras, India) (Ed.), Vedic River Sarasvati and Hindu civilization. pp. 398.

van Driem, G., 2001. Languages of the Himalayas: an ethnolinguistic handbook of the greater Himalayan region containing an introduction to the symbiotic theory of language. Brill, Leiden, The Netherlands.

Vaughan, D., Lu, B., Tomooka, N., 2008. The evolving story of rice evolution. Plant Science 174, 394–408.

Vijh, R.K. (Ed.), 2000. Domestic animal diversity.

Vitte, C., Ishii, T., Lamy, F., Brar, D., Panaud, O., 2004. Genomic paleontology provides evidence for two distinct origins of Asian rice (Oryza sativa L.). Molecular Genetics and Genomics 272, 504–511.

Vivekananda, S., 1947. Caste, culture, and socialism Advaita Ashrama.

Vivekananda, S., 1970. The complete works of Swami Vivekananda Advaita Ashrama.

Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K., 2006. A map of recent positive selection in the human genome. PLoS Biol 4, e72.

Wall, J.D., Kim, S.K., 2007. Inconsistencies in Neanderthal genomic DNA sequences. PLoS Genet 3, 1862–1866.

Wang, G., Mackill, D., Bonman, J., McCouch, S., Champoux, M., Nelson, R., 1994. RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics 136, 1421.

Watkins, W.S., Bamshad, M., Dixon, M.E., Bhaskara Rao, B., Naidu, J.M., Reddy, P.G., Prasad, B.V., Das, P.K., Reddy, P.C., Gai, P.B., Bhanu, A., Kusuma, Y.S., Lum, J.K., Fischer, P., Jorde, L.B., 1999. Multiple origins of the mtDNA 9-bp deletion in populations of South India. Am J Phys Anthropol 109, 147–158.

Watkins, W.S., Prasad, B.V.R., Naidu, J.M., Rao, B.B., Bhanu, B.A., Ramachandran, B., Das, P.K., Gai, P.B., Reddy, P.C., Reddy, P.G., Sethuraman, M., Bamshad, M.J., Jorde, L.B., 2005. Diversity and divergence among the tribal populations of India. Ann Hum Genet 69, 680–692.

Watson, E., Forster, P., Richards, M., Bandelt, H.J., 1997. Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61, 691–704.

Weber, S., 1998. Out of Africa: The Initial Impact of Millets in South Asia. Curr Anthropol 39, 267–274.

Weiss, K.M., Long, J.C., 2009. Non-Darwinian estimation: my ancestors, my genes' ancestors. Genome Res 19, 703–710.

Wells, R.S., Yuldasheva, N., Ruzibakiev, R., Underhill, P.A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., Balakrishnan, K., Read, M., Pearson, N.M., Zerjal, T., Webster, M.T., Zholoshvili, I., Jamarjashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zalloua, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A., Bodmer, W.F., 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA 98, 10244–10249.

Westgate, J., Shane, P., Pearce, N., Perkins, W., Korisettar, R., Chesner, C., Williams, M., Acharyya, S., 1998. All Toba tephra occurrences across peninsular India belong to the 75 000 yr BP eruption. Quaternary Research 50, 107–112.

Wheeler, M., 1979. Harappan Chronology and the Rig Veda. Ancient Cities of the Indus.

White, D.J., Wolff, J.N., Pierson, M., Gemmell, N.J., 2008. Revealing the hidden complexities of mtDNA inheritance. Mol Ecol 17, 4925–4942.

White, T.D., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G.D., Suwa, G., Howell, F.C., 2003. Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature 423, 742–77.

Whitehouse, P., 1997. “The External Relationships of the Nihali and Kusunda Languages.”. Mother Tongue 3, 4–44.

Whitehouse, P., Usher, T., Ruhlen, M., Wang, W.S.Y., 2004. Kusunda: an Indo-Pacific language in Nepal. Proc Natl Acad Sci U S A 101, 5692–5695.

Wiik, K., 2002. Eurooppalaisten juuret Atena, Jyväskylä.

Wilson, M., Stoneking, M., Holland, M., DiZinno, J., Budowle, B., 1993. Guidelines for the use of mitochondrial DNA sequencing in forensic science. Crime Lab Digest 20, 68–77.

Witzel, M., 1999. Substrate languages in old Indo-Aryan. Electronic Journal of Vedic Studies. 5–1, 1–67.

Witzel, M., 2005. Central Asian roots and acculturation in Indian subcontinent: linguistic and archaeological evidence from Western Central Asia, the Hindukush and northwestern Indian subcontinent for early Indo-Aryan language and religion, in: Osada, T. (Ed.), Liguistics, Archaeology and the Human Past. Research Institute for Humanity and Nature, Koyto, Koyto, Japan, pp. 87–211.

Witzel, M., 2007. The Languages of Harappa, in J. Knoyer (Ed.), Indus civilization.

Wiuf, C., Hein, J., 1997. On the number of ancestors to a DNA sequence. Genetics 147, 1459–1468.

Xing, J., Watkins, W.S., Witherspoon, D.J., Zhang, Y., Guthery, S.L., Thara, R., Mowry, B.J., Bulayeva, K., Weiss, R.B., Jorde, L.B., 2009. Fine-scaled human genetic structure revealed by SNP microarrays. Genome Res 19, 815–825.

Xue, Y., Zerjal, T., Bao, W., Zhu, S., Shu, Q., Xu, J., Du, R., Fu, S., Li, P., Hurles, M.E., Yang, H., Tyler-Smith, C., 2006. Male demography in East Asia: a north-south contrast in human population expansion times. Genetics 172, 2431–2439.

Xue, Y., Wang, Q., Long, Q., Ng, B.L., Swerdlow, H., Burton, J., Skuce, C., Taylor, R., Abdellah, Z., Zhao, Y., Asan, MacArthur, D.G., Quail, M.A., Carter, N.P., Yang, H., Tyler-Smith, C., 2009a. Human Y-chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. Curr Biol 19, 1453–1457.

Xue, Y., Zhang, X., Huang, N., Daly, A., Gillson, C.J., Macarthur, D.G., Yngvadottir, B., Nica, A.C., Woodwark, C., Chen, Y., Conrad, D.F., Ayub, Q., Mehdi, S.Q., Li, P., Tyler-Smith, C., 2009b. Population differentiation as an indicator of recent positive selection in humans: an empirical evaluation. Genetics 183, 1065–1077.

Yao, Y.G., Watkins, W.S., Zhang, Y.P., 2000. Evolutionary history of the mtDNA 9-bp deletion in Chinese populations and its relevance to the peopling of East and Southeast Asia. Hum Genet 107, 504–12.

Yi, S., Ellsworth, D.L., Li, W.H., 2002. Slow molecular clocks in Old World monkeys, apes, and humans. Mol Biol Evol 19, 2191–2198.

Yngvadottir, B., 2007. Insights into modern disease from our distant evolutionary past. European Journal of Human Genetics 15, 603–606.

Zerjal, T., Pandya, A., Thangaraj, K., Ling, E.Y.S., Kearley, J., Bertoneri, S., Paracchini, S., Singh, L., Tyler-Smith, C., 2007. Y-chromosomal insights into the genetic impact of the caste system in India. Hum Genet 121, 137–144.

Zhivotovsky, L.A., Underhill, P.A., Cinnioğlu, C., Kayser, M., Morar, B., Kivisild, T., Scozzari, R., Cruciani, F., Destro-Bisol, G., Spedini, G., Chambers, G.K., Herrera, R.J., Yong, K.K., Gresham, D., Tournev, I., Feldman, M.W., Kalaydjieva, L., 2004. The effective mutation rate at Y-chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet 74, 50–61.

Zide, N.H., Barker, M.E. (Ed.), 1966. Studies in comparative Austroasiatic linguistics.

SUMMARY IN ESTONIAN

Selguse mõttes määratleme Lõuna-Aasiat traditsiooniliselt kui regiooni, mis kaasaegsest poliitilisest kaardist lähtuvalt haarab endasse India, Pakistani ja Bangladeshi. Lõuna-Aasia rahvaste geneetiline ajalugu on olnud pikki aastaid inimkonna demograafilist ajalugu uurivate teadlaste oluliseks huviobjektiks. Selleks on mitmeid põhjusi. Inimkonna kui terviku seisukohast on oluline muidugi see, et nimetatud piirkonnas elab veidi üle pooleteise miljardi inimese, ehk 22% maailma rahvastikust – mitmekordselt enam, kui suvalises teises piirkonnas väljapool Aasiat. Inimkonna demograafilise ajaloo seisukohast võib pidada väga oluliseks ka asjaolu, et pikal perioodil peale Euraasia koloniseerimise algust anatoomiliselt kaasaegse inimese poolt, oli mitmete arvutuste põhjal Lõuna-Aasia suurima elanikkonnaga piirkonnaks maakeral. Veelgi enam – kuigi Lõuna-Aasia paleoliitilikumi arheoloogia ei ole rikkalik, peavad paljud uurijad just Lõuna-Aasiat kõige olulisemaks migratsiooniteeks, mille kaudu inimkond jõudis edasi Hiinasse ning Siberisse, ja sealt edasi Ameerikasse ning paralleelselt sellele Kagu-Aasiasse ning Austraaliasse.

Kuid lisaks neile väga üldistele kaalutlustele, pakub Lõuna-Aasia ja eriti India geneetilise struturereerituse uurimine suurt huvi ka põhjusel, et siin on tegemist erakordselt keeruka ja ajalooliselt ammu tagasi väljakujunenud sotsiaalse struktuuriga regiooniga, Kuigi kastisüsteem on tänapäeval lagunemas, eriti linnastumise tulemusel, on tema mõju endiselt tugev ka praegu, väljendudes kauaaegsetes tõketes geenide vabale liikumisele, endogaamiana kitsamas ja laiemas tähenduses. Indias kõneldakse paljusid keeli, kuid keelkondade mõttes on valdavaks indo-euroopa põhjapoolsetel aladel ja draviidi keeled lõuna-poolsetel. Arvukalt elab Indias ka austro-aasia keelkonna munda, samuti tiibeti-birma keelkonna keelte kasutajaid. Seetõttu ongi paljude geneetikute püüdluseks olnud selgitada erinevusi eri keelegruppide kasutajate geneetikas. Seal-juures on eriti suurt tähelepanu pööratud sotsiaalsete kastide geneetilisele erinevusele ja üldkultuuriliselt maailma teadlaskonnale suurt huvi pakkuvate intrigeerivatele probleemidele, nagu seda in näiteks idee sellest, et indo-euroopa keeled ja nende kandjad (st. geenifond) jõudis Indiasse suhteliselt hiljuti, kolme-nelja tuhande aasta eest Kesk-Aasiast ja langeb kokku "veedade ajastu" algusega. Ei ole selgust ka draviidi keelte "saabumise" kohta Indiasse ja selle seosesse elanikkonna geneetilise ajalooga. Kui nii, siis kes olid pre-indo-aaria ning pre-draviidi asukad geneetiliselt? Kokkuvõtteks võib öelda, et tänu pika-ajalisele, kümneid aastatuhandeid väldanud asustusele, aastatuhandete sügavusele tsivilisatsioonile, erakordselt keerulisele sotsiaalsele struktuurile ja ebaselgele keelelisele ajaloole, millele lisandub pooleteisemiljardine rahvastik, on Lõuna-Aasia väljakutseks kõigile, kes soovivad uurida inmkonna geneetilise struktureerituse kujunemist ja tänapäeva selle keerukuses. Need on üldist laadi kaalutlused, mis teevad Lõuna Aasia rahvaste geneetika uurimise oluliseks.

Kuid on ka otseseid põhjusi, miks dissertatsioon on pühendatud India geneetikale. Evolutsioonilise bioloogia õppetool ja Eesti Biokeskus on olnud tosina aasta jooksul kohaks, kus on muude teemadega paralleelslt uuritud ka

Lõuna-Aasia geneetikat. Hyderabadi Raku- ja Molekulaarbioloogia Keskus Indias on keskseks India teaduskeskuseks India rahvaste populatsioonigeneetika uurimisel. Seal algas minu teaduslik tegevus. Nende kahe laboratooriumi eelnev ja tänini jätkuv koostöö tegi võimalikuks formuleerida mitmeid uusi teaduslikke küsimusi vastvalt eelnevalt omandatud teadmistele ja lahendamist ootavatele probleemidele. Neid küsimusi on muidugi märksa enam, kui minu dissertatsioon katta suudb, kuid seletuste otsimine alamosale neist ongi olnud käesoleva väitekirja ja selle aluseks olevate teaduslike artiklite sisuks. Samuti ka rea teiste avaldatud ja avaldamisele suunatud uuringute valdavaks sisuks, mille kaasautoriks ma olen olnud.

Väitekirja konkreetsed eesmärgid (täpsemini – väitekirja sisuks valitud konkreetsete artiklite temaatika) baseeruvad varasemal teadmisel, et India rahvastiku emaliinis päranduva mitokondriaalse DNA varieeruvus on ulatuslikut dikteeritud makro-haplogruppide M ja R variantide poolt. Meie ja teiste laborite senise analüüsi põhjal võis järeldada, et India puhul on lõviosas tegemist variantidega, mis ei esine Ida-Aasias, kuid uuringute sügavus (mtDNA sekveneerimise ulatus) ei võimaldanud seda paljudel juhtudel siiski kindlalt väita. Samas on just teadmine sellest, kas vastav geeniliinide klaster pärineb vahetult hg M või siis hg R fülogeneetilistest alguspunktidest (*ancestral states*), suure väärtusega mõistmaks Euraasia asustamist ajalis-ruumilises dimensioonis.

Oluliseks konkreetseks eesmärgiks oli uurida India praeguste ja "äsjaste" hõimude (nn *sceduled castes*, mille täpsema seletuse leiab kirjanduse ülevaatest) emaliini-järgset geneetikat. Selle küsimuse selgitamise taga on tegelikult mitu olulist alamküsimust. Nimelt on mitmed antropoloogid oletanud, et need rahvagrupid, alles hiljuti paljuski kütid-korilased, on "tõelised" India ürgasukad, samas kui indo-euroopa ja draviidi keeli rääkivad Lõuna-Aasia elanikud on Indiasse jõudnud palju hilisemate rännete tulemusena.

Teiseks probleemide grupiks oli spetsiifiliselt munda-keelsete rahvagruppide keele ja geenide vahekord. Munda keeled on alamhulk austro-aasia keeltest, mida kõneldake laias ulatuses mitmel pool Kagu-Aasias, kuid ka India ida-poolseis provintsides. Meid huvitas nende gruppide geneetiline identiteet. Üheks võimaluseks valisime välja neil sageli esineva mtDNA haplogrupi R7 ülatuslikuma resekveneerimise nii munda-keelsetel indialastel, kui ka draviidi ja indo-euroopa keeli rääkivail indialastel.

Töö eesmärkide hulka kuulus ka Andamani saarestiku elanike geneetika uurimine, sest et selle saarerühma päriselanike päritolu on olnud väga mitmesuguste oletuste objektiks. Sealjuures põhjusel, et oma väljanägemiselt kuuluvad nad klassikalise antropoloogia jaotuse kohaselt negriitode hulka – nad on väga tumeda nahaga ja lokkisjuukselised, valdavalt lühikest kasvu. Eelnevalt oli leitud, et Andamani negriitodel esinev mtDNA variant on kauges fülogeneetilises suguluses Indias esinevate haplogruppidega. Oletuste kinnitamiseks (või ümberlükkamiseks) viisime läbi uurimuse, mis haaras erinevaid Andamani saari.

Töös saadud tulemuste alusel formuleeriti alljärgnevad põhilised järeldused:

1. India mtDNA variandid sisaldavad endas arvukalt fülogeneetiliselt sügavaid harusid, mis tõendavad vastavate klaadide *in situ* teket Lõuna-Aasias. Arvukate autohtoonsete hg M ja hg R geeniliinide esinemine selles regioonis, millel praktiliselt ei ole leitud alamharusid mujal Aasias ega ka Okeaanias, on lisakinnituseks arvamusele, et anatoomiliselt kaasaegse inimkonna põhiline väljaränd Aafrikast oli ühekordne.
2. Kooskõlas Andamani põliselanike geograafilise ja kultuurilise isolatsiooniga me näitasime, et hg M31 jaguneb seal kaheks alamhulgaks, milledest M31a1 esineb Onge ja Jarwa (*pro* Jarawa) elanikel, kuid M31a2 vaid Suurtel Andamanidel.
3. Uurides Põhja-India hindi-keelset Mushari rahva geneetikat leidsime, et nende puhul on väga tõenäoliselt tegemist Lõuna-Aasiale iseloomuliku protsessiga, kus on toimunud (kiire) keeleline muutus ilma, et sellest haaratud populatsiooni genofond oleks muutunud. Üldistatult võib märkida, et niisugune fenomen (*language shift*), mida ennemalt peeti pigem haruldaseks, näib nüüd, tänu geneetika meetodite täpsustumisele, olevat väga üldiseks fenomeniks nii Indias kui üle kogu maailma.
4. Toetudes Lõuna-Aasia autohtoonsete emajärgsete geeniliinide (mtDNA) olemasolule ja nende hargnemise ajalisele sügavusele, mis on ilmselt paljukordset varasemad praegu Lõuna-Aasias levinenud indo-euroopa, draviidi ja austro-aasia keelkondade tekkest Euraasias, teeme järelduse, et India geneetilise varieeruvuse maastik on seletetav esmajoones geograafia, mitte keel(te)ga. Kindlasti on ka viimaste varieeruvus oma mõju avaldanud, esmajoones lokaalselt ning käsitletuna koos sotsiaalse kihistumisega, kuid need mõjud on olnud hilisemad. Nende mõjude ulatus ja kvantitatiivne ning kvalitatiivne hindamine on meie praeguste ja tulevaste uuringute huvitavaks eesmärgiks.

ACKNOWLEDGEMENTS

I am indebted to my supervisor Dr. Toomas Kivisild for his able guidance, constant encouragement, freedom of thought and constructive suggestions which helped me to learn the A, B, C of human population genetics. His ideas and concepts have had a remarkable influence throughout my career. His research on Indian prehistory was a source of inspiration for me to switch from Drosophila genetics to human population genetics. Working under him has been a fabulous experience at scientific as well as personal level and I shall forever treasure this association.

I am grateful to Professor Richard Villems not only for the constructive advice he gave me at all stages of my thesis but also for building a scientific culture and temperament in my life. Working with him has been a great experience and it is an honor for me to be a part of his workgroup. I sincerely thank my co-supervisor Dr. Ene Metspalu for her suggestions and cordial help. I also express my deep sense of gratitude to Dr. Mait Metspalu for his keen interest and unhesitant support. I take the opportunity to thank Dr. Siiri Rootsi for her suggestions and participation in fruitful discussions related with Y-chromosomal work.

I wish to thank Anu, Phillip and Juan for their help during the early stages of my PhD tenure. I am thankful to Maere, Erika, Eva-Liis, Kristiina, Mari, Helle-Viivi, Georgi, Erik and Hovo for creating a congenial atmosphere. My colleagues and friends Monika and Erwan deserves special thanks for providing a good company and being extremely supportive. I express heartfelt thanks to Jüri, Ille, Tuuli and Jaan for their technical assistance and non-stop supply of the reagents for my experiments. I also acknowledge Jüri for showing me Estonian nature (wild pigs, seals and birds) by live streaming.☺

I express my gratitude to my previous supervisors, Dr. K. Thangaraj and Dr. Lalji Singh, CCMB, Hyderabad, who taught me the basics and always inspired me to think better. It was a great pleasure working under them in CCMB Hyderabad, India. I am extremely thankful to all my international collaborators and co-authors especially Prof. Antonio Torroni, Dr. Chris-Tyler Smith, Dr. Marta Lahr, Prof. Rob Foley, Prof. Martin Richards, Dr. Peter Underhill for their constructive inputs and suggestions in the manuscript preparation. I am grateful to Prof. Andres Metspalu and Dr. Urmas Saarma for their helpful suggestions to improve the quality of dissertation being expert reviewers.

I was fortunate to have very supportive international friends in the institute and I thank them, especially Lena, Mannis, Nina (and her C*), Irene, Anna, Marian, Sardana, Ildus, Bayazit, Tonka, Lejla and Merisa. Many thanks to Feddy for proof reading the Nahali's chapter. I would like to acknowledge the support extended by my friends especially Vijay, Ajai, Deoji, Shahrukhji and all Indian friends in Tartu. I would also like to acknowledge Indian Cultural Association of Estonia and its President Mr. Naresh for creating a homely atmosphere.

The financial assistance in the form of PhD Fellowship by the University of Tartu and the Estonian Biocentre are duly acknowledged. I take this opportunity to acknowledge the International Student Office staff, especially Ms. Ülle Tensing and Piret for their timely support and enthusiastic help whenever needed. I specially owe great obligations to all those people who donated their sample for study and also those who helped in sampling.

I am highly grateful to my parents and sisters for being a constant source of encouragement and strength. I fall short of words in expressing their love, affection, and belief in my dreams and ceaseless support towards achieving my aspirations. I am obliged to my grandparents who taught me the essence of life. Their influence is all over these pages and all over my life. I express my heartiest gratitude to my wife, Chandana for being a source of inexhaustible encouragement, unconditional love and inspiration.

Gyaneshwer Chaubey

PUBLICATIONS

CURRICULUM VITAE

Personal Details

Name	Gyaneshwer Chaubey
Date of Birth	2nd September 1979
Mother's Name	Smt. Vidya Chaubey
Father's Name	Sri. S. N. Chaubey
Permanent Address	Vill.- Sungulpur, P.O.- Chaubeypur Distt.-Varanasi, Uttar Pradesh- 221104 India
Address for Correspondence	Institute of Molecular and Cell Biology, University of Tartu and Estonian Biocentre Riia 23, Tartu Estonia, 51010
Phone (Office)	+372 737 5005
E-mail	gyanc@ebc.ee

Academic Background

From 2005	PhD student. Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu, Estonia.
2003–2005	Project Assistant in the "Human Genome Diversity" project under the supervision of Dr. K. Thangaraj and Dr. Lalji Singh, Centre for Cellular and Molecular Biology (CCMB), Hyderabad, India.
2001–2003	Master of Science student (M. Sc.) in Biotechnology. V.B.S. Purvanchal University Jaunpur, Uttar Pradesh, India.
2003	MSc. Dissertation from Centre for Cellular and Molecular Biology Hyderabad, India. Title: "Y chromosome and mtDNA polymorphism among Brahmin Caste of Uttar Pradesh, Santhal tribe of Jharkhand and Kanwar Tribe of Chhattisgarh" under supervision of Dr. K. Thangaraj and Dr. Lalji Singh.
2002	MSc. Summer Training from Cytogenetics Laboratory, Banaras Hindu University (BHU) Varanasi, India. Title: "Effect of Hsr-ω mutation on distribution of *Hrb57A* and Actin in developing egg chambers of *Drosophila melanogaster*" in the supervision of Prof. S.C. Lakhotia, (May 2002–September 2002).

Distinctions and Awards

(i) Qualified GATE 2003 and CSIR (Council of Scientific and Industrial Research) National Eligibility Test (NET) for Lectureship/Assistant Professorship in Life Sciences, twice in June 2003 and December 2003.
(ii) Won the "Travel Award" to participate in HUGO conference, Hyderabad-October 2008, and SMBE conference Lyon July-2010.

Papers and Poster Presentations in international meetings

(i) Presented a paper entitled "Bidirectional gene flow between Austroasiatic speakers of India and Southeast Asia" in SMBE conference Lyon, France, July 2010.
(ii) Presented a poster entitled "The Genetic origin of Austroasiatic speakers" in HGV conference Tallinn, Estonia, Sept. 2009.
(iii) Presented a paper entitled "Genetics of 'Gotra' and 'Marriage pattern' among Brahmins" in ISHG conference New Delhi-March 2009.
(iv) Presented a paper "Population Genetics of Brahmins and 'Gotra' system in India" in UKIERI conference LCHES, Cambridge Unuversity, UK Nov.2008.
(v) Presented a poster "Intricate analysis of Austro-Asiatic speaking populations portrays their origin to Southeast Asia" in HUGO conference, Hyderabad-October 2008.
(vi) Presented a paper entitled "Phylogeography of Haplogroup R7" in Human Evolution and Disease, EMBO India CCMB, Hyderabad conference Dec 2006.

Publications:

Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, Parik J, Rootsi S, **Chaubey G**, Kutuev I, Yudkovsky G, Khusnutdinova EK, Balanovsky O, Semino O, Pereira L, Comas D, Gurwitz D, Bonne-Tamir B, Parfitt T, Hammer MF, Skorecki K, Villems R. 2010. The genome-wide structure of the Jewish people. ***Nature*** 466:238–242.

Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovsky LA, King RJ, Lin AA, Chow CET, Semino O, Battaglia V, Kutuev I, Järve M, **Chaubey G,** Ayub Q, Mohyuddin A, Mehdi SQ, Sengupta S, Rogaev EI, Khusnutdinova EK, Pshenichnov A, Balanovsky O, Balanovska E, Jeran N, Augustin DH, Baldovic M, Herrera RJ, Thangaraj K, Singh V, Singh L, Majumder P, Rudan P, Primorac D, Villems R, Kivisild T. 2010. Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a***. Eur J Hum Genet*** 18:479–484.

Eaaswarkhanth M, Haque I, Ravesh Z, Romero IG, Meganathan PR, Dubey B, Khan FA, **Chaubey G,** Kivisild T, Tyler-Smith C, Singh L, Thangaraj K. 2010. Traces of sub-Saharan and Middle Eastern lineages in Indian Muslim populations. ***Eur J Hum Genet*** 18:354–363.

Petraglia M, Clarkson C, Boivin N, Haslam M, Korisettar R, **Chaubey G,** Ditchfield P, Fuller D, James H, Jones S, Kivisild T, Koshy J, Lahr MM, Metspalu M, Roberts R, Arnold L. 2009. Population increase and environmental deterioration correspond with microlithic innovations in South Asia ca. 35,000 years ago. ***Proc Natl Acad Sci U S A*** 106:12261–12266.

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*** 8:227.

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

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*** 66:1–9.

Chaubey G, Metspalu M, Kivisild T, Villems R. 2007. Peopling of South Asia: investigating the caste-tribe continuum in India. ***Bioessay***s 29:91–100.

Chaubey G, Metspalu M, Villems R, Kivisild T. 2007. Reply to winters. ***Bioessays*** 29:499.

Thangaraj K, **Chaubey G,** Singh VK, Reddy AG, Chauhan P, Malvee R, Pavate PP, Singh L. 2007. Y-chromosomal STR haplotypes in two endogamous tribal populations of Karnataka, India. ***J Forensic Sc***i 52:751–753.

Thangaraj K, **Chaubey G,** Singh VK, Reddy AG, Pavate PP, Singh L. 2006. Genetic profile of nine autosomal STR loci among Halakki and Kunabhi populations of Karnataka, India. ***J Forensic Sci*** 51:190–192.

Thangaraj K, **Chaubey G,** Reddy AG, Singh VK, Singh L. 2007. Autosomal STR data on the enigmatic Andaman Islanders. ***Forensic Sci Int*** 169:247–251.

Thangaraj K, **Chaubey G,** Reddy AG, Singh VK, Singh L. 2006. Unique origin of Andaman Islanders: insight from autosomal loci. ***J Hum Genet*** 51:800–804.

Thanseem I, Thangaraj K, **Chaubey G,** Singh VK, Bhaskar LVKS, Reddy BM, Reddy AG, Singh L. 2006. Genetic affinities among the lower castes and

tribal groups of India: inference from Y chromosome and mitochondrial DNA. ***BMC Genet*** 7:42.

Thangaraj K, **Chaubey G,** Singh VK, Vanniarajan A, Thanseem I, Reddy AG, Singh L. 2006. In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. ***BMC Genomics*** 7:151.

Pandu G, Gandhi KPC, Sharma JD, **Chaubey G,** Thangaraj K. 2006. Genetic profile of nine STR loci among Goud and Padmashali populations of Andhra Pradesh, India***. Forensic Sci Int*** 157:201–205.

Thangaraj K, **Chaubey G,** Kivisild T, Reddy A, Singh V, Rasalkar A, Singh L. 2006. Response to Comment on "Reconstructing the Origin of Andaman Islanders***". Science*** 311:470b.

Thangaraj K, **Chaubey G,** Kivisild T, Reddy A, Singh V, Rasalkar A, Singh L. 2005. Tracing modern human origins – Response. ***Science*** 309:1996–1997.

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*** 116: 507–517.

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

ELULOOKIRJELDUS

Isikuandmed

Nimi:	Gyaneshwer Chaubey
Sünniaeg:	2. september 1979
Aadress kirjavahetuseks:	TÜMRI ja Eesti Biokeskus, Riia 23, 51010 Tartu, Eesti
Telefon tööl:	+372 737 5005
e-post:	gyanc@ebc.ee

Akadeemiline elukäik

alates 2005	Doktoriõpe Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituudi evolutsioonilise bioloogia õppetoolis
2003–2005	Projektiassistent projektis "Inimgenoomi varieeruvus" Dr K. Thangaraj ja Dr. Lalji Singi juhendamisel, Raku- ja Molekulaarbioloogia Keskus (CCMB), Hyderabad, India
2001–2003	Magistratuur biotehnoloogia erialal, V.B.S. Purvanchali Ülikool, Jaunpur, Uttar Pradesh, India
2003	Magistriväitekiri Raku- ja Molekulaarbioloogia Keskusest (CCMB) teemal "Y-kromosoomi ja mtDNA polümorfism Uttar Pradeshi brahmiinidel, Santali hõimul Jharkhandis ja Kanwari hõimul Chhattisgarhis" Dr. K. Thangaraji ja Dr. Lallji Singhi juhendamisel.
2002	Suvine magistrikursus Banarasi Hindu Ülikooli (BHU) tsütogeneetika laboratooriumis, Varanasi, India. Teema: "Hrs-ω +mutatsiooni mõju *Hrb57A* ja aktiini jaotumisele *Drosophila melanogasteri* arenevas munakambris" prof. S.C. Lakhota juhendamisel.

Tunnustused ja auhinnad

(i) Rahvuslik kvalifikatsioonitunnistus (NET) lektori/assistendi ametikohale eluteadustes Teaduse ja Tööstusuuringute Nõukogult), 2003. a. juuni ning detsember.

(ii) "Reisipreemia" osavõtuks HUGO 2008. a. kongressist Hyderabadis ja SMBE konverentsist Lyonis 2010. a

Suulised ja posteriettekanded rahvusvahelistel konverentsidel

(i) Suuline ettekanne “Bidirectional gene flow between Austroasiatic speakers of India and Southeast Asia", SMBE konverents, Lyon 2010.
(ii) Poster “The Genetic origin of Austroasiatic speakers”, Human Genome Variation” konverents, Tallinn 2009.
(iii) Suuline ettekanne “Gotrate süsteemi geneetika ja brahmiinide abielumustrid”, ISHG konverent, New Delhi, 2009.
(iv) Suuline ettekanne “Brahmiinide populatsioonigeneetika ja gotrate süsteem Indias”, UKIERI konverents, LCHES, Cambridge Ülikool, UK, 2008.
(v) Poster “Austroaasia keeli rääkivate populatsioonide süvaanalüüs osutab nende päritolule Kagu-Aasiast”, HUGO konverents, Hyderabad, India.
(vi) Suuline ettekanne “Haplogrupp R7 fülogeograafia”, EMBO India konverents, Hyderabad, India, 2006.

Teaduslikud kirjutised

Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, Parik J, Rootsi S, **Chaubey G**, Kutuev I, Yudkovsky G, Khusnutdinova EK, Balanovsky O, Semino O, Pereira L, Comas D, Gurwitz D, Bonne-Tamir B, Parfitt T, Hammer MF, Skorecki K, Villems R. 2010. The genome-wide structure of the Jewish people. ***Nature*** 466:238–242.

Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovsky LA, King RJ, Lin AA, Chow CET, Semino O, Battaglia V, Kutuev I, Järve M, **Chaubey G,** Ayub Q, Mohyuddin A, Mehdi SQ, Sengupta S, Rogaev EI, Khusnutdinova EK, Pshenichnov A, Balanovsky O, Balanovska E, Jeran N, Augustin DH, Baldovic M, Herrera RJ, Thangaraj K, Singh V, Singh L, Majumder P, Rudan P, Primorac D, Villems R, Kivisild T. 2010. Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a***. Eur J Hum Genet*** 18:479–484.

Eaaswarkhanth M, Haque I, Ravesh Z, Romero IG, Meganathan PR, Dubey B, Khan FA, **Chaubey G,** Kivisild T, Tyler-Smith C, Singh L, Thangaraj K. 2010. Traces of sub-Saharan and Middle Eastern lineages in Indian Muslim populations***. Eur J Hum Genet*** 18:354–363.

Petraglia M, Clarkson C, Boivin N, Haslam M, Korisettar R, **Chaubey G,** Ditchfield P, Fuller D, James H, Jones S, Kivisild T, Koshy J, Lahr MM, Metspalu M, Roberts R, Arnold L. 2009. Population increase and environmental deterioration correspond with microlithic innovations in South Asia ca. 35,000 years ago***. Proc Natl Acad Sci U S A*** 106:12261–12266.

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*** 8:227.

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

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*** 66:1–9.

Chaubey G, Metspalu M, Kivisild T, Villems R. 2007. Peopling of South Asia: investigating the caste-tribe continuum in India. ***Bioessay***s 29:91–100.

Chaubey G, Metspalu M, Villems R, Kivisild T. 2007. Reply to winters. ***Bio-essays*** 29:499.

Thangaraj K, **Chaubey G,** Singh VK, Reddy AG, Chauhan P, Malvee R, Pavate PP, Singh L. 2007. Y-chromosomal STR haplotypes in two endogamous tribal populations of Karnataka, India***. J Forensic Sc***i 52:751–753.

Thangaraj K, **Chaubey G,** Singh VK, Reddy AG, Pavate PP, Singh L. 2006. Genetic profile of nine autosomal STR loci among Halakki and Kunabhi populations of Karnataka, India***. J Forensic Sci*** 51:190–192.

Thangaraj K, **Chaubey G,** Reddy AG, Singh VK, Singh L. 2007. Autosomal STR data on the enigmatic Andaman Islanders***. Forensic Sci Int*** 169:247–251.

Thangaraj K, **Chaubey G,** Reddy AG, Singh VK, Singh L. 2006. Unique origin of Andaman Islanders: insight from autosomal loci***. J Hum Genet*** 51:800–804.

Thanseem I, Thangaraj K, **Chaubey G,** Singh VK, Bhaskar LVKS, Reddy BM, Reddy AG, Singh L. 2006. Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. ***BMC Genet*** 7:42.

Thangaraj K, **Chaubey G,** Singh VK, Vanniarajan A, Thanseem I, Reddy AG, Singh L. 2006. In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. ***BMC Genomics*** 7:151.

Pandu G, Gandhi KPC, Sharma JD, **Chaubey G,** Thangaraj K. 2006. Genetic profile of nine STR loci among Goud and Padmashali populations of Andhra Pradesh, India***. Forensic Sci Int*** 157:201–205.

Thangaraj K, **Chaubey G,** Kivisild T, Reddy A, Singh V, Rasalkar A, Singh L. 2006. Response to Comment on "Reconstructing the Origin of Andaman Islanders***". Science*** 311:470b.

Thangaraj K, **Chaubey G,** Kivisild T, Reddy A, Singh V, Rasalkar A, Singh L. 2005. Tracing modern human origins – Response. ***Science*** 309:1996–1997.

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*** 116: 507–517.

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

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 crassland 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.
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, 2008, 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 Kõster.** 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.