DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
114

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
114

THROUGH THE COURSE OF PREHISTORY IN INDIA: TRACING THE mtDNA TRAIL

MAIT METSPALU

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 14.11.2005, by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Opponent: Antonio Salas Ellacuriaga, PhD, Institute of Legal Medicine, Faculty of Medicine University of Santiago de Compostela

Commencement: Room No 217, Riia 23, Tartu, on December 19th, at 14.00

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

ISSN 1024–6479
ISBN 9949–11–201–X (trükis)
ISBN 9949–11–202–8 (PDF)

Autoriõigus Mait Metspalu, 2005

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 578

CONTENTS

LIST OF ORIGINAL PUBLICATIONS 7

ABBREVIATIONS 8

1. INTRODUCTION 9

2. LITERATURE OVERVIEW 10
   2.1. Some recent developments regarding the properties of mtDNA in respect to phylogenetic studies 10
      2.1.1. Mitochondrial Steve – a gloomy story of a substitute player who somehow never really got on the field thus far. 12
      2.1.2. Positive selection towards heat production through less-efficient OXPHOS in the freezing North? Problematic landing for a nice idea. 14
   2.2. The pioneer settlement of modern humans in Asia 16
      2.2.1 Introduction 16
      2.2.2. Palaeoclimatological context 17
      2.2.3. Archaeological and palaeontological evidence of the peopling of Asia by AMH 18
      2.2.4. How to reconstruct "pioneer settlement" from extant MtDNA diversity 20
      2.2.5. The peopling of Asia as seen through the lens of mtDNA diversity 22
      2.2.6. A route through Northern Asia? 26
      2.2.7. Conclusion 29
   2.3. India: Some general considerations 30
      2.3.2. Palaeontological and archaeological data 33
      2.3.3. Linguistic data 37
      2.3.4. Data obtained from studies using "classical" markers 40

3. AIMS OF THE PRESENT STUDY 42

4. RESULTS AND DISCUSSION 43
   4.1. A deep link between South and West Eurasian maternal lineages (Ref I) 43
   4.2. The caste and tribal populations of India share a common ancestry in Late Pleistocene (Ref II) 43
   4.3. Post Pleistocene spread of languages and farming into South Asia shaped the pre-existing mtDNA pool only marginally (Ref III) 45

4.4. The extant mtDNA pool of South Asia was inoculated during the initial peopling of the continent and has not been replaced since (Ref IV) 46

5. CONCLUSIONS 48

REFERENCES 49

KOKKUVÕTE 59

ACKNOWLEDGEMENTS 60

PUBLICATIONS 61

LIST OF ORIGINAL PUBLICATIONS

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

I. Kivisild T, Bamshad M, Kaldma K, **Metspalu M,** Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R (1999). Deep common ancestry of Indian and western Eurasian mtDNA lineages. Current Biology 9: 1331–1334.

II. Kivisild T, Rootsi S, **Metspalu M**, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk H-V, Stepanov V, Gölge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R (2003a) *The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations.* American Journal of Human Genetics 72:313–332.

III. 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 (eds) Examining the farming/language dispersal hypothesis. The McDonald Institute for Archaeological Research, Cambridge, pp 215–222

IV. **Metspalu M**, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, 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 Genetics 5:26

My contribution to the articles referred in the current thesis is as follows:

Ref. **I** – a) participated in performing the experiments: b) assisted in the analysis of the data and; c) in the preparation of the manuscript;

Ref. **II** – a) participated in performing the experiments: b) assisted in the analysis of the data and; c) in the preparation of the manuscript;

Ref. **III** – a) participated in performing the experiments b) assisted in the analysis of the data and; c) in the preparation of the manuscript;

Ref. **IV** – a) had a key role in the design of the study; b) performed most of the mtDNA experiments c) performed the statistical and phylogeographical analysis of mtDNA; d) wrote the paper;

ABBREVIATIONS

AMH	anatomically modern human(s)
bp/kbp	base pair/thousand (kilo) base pairs
hg(s)	haplogroup(s)
HVS-I/HVS-II	the first/second hypervariable segment of mtDNA
LD	linkage disequilibrium
LGM	the Last Glacial Maximum
MRCA	the most recent common ancestor
MTATP6	Subunit 6 of ATP sythase (complex V)
MTCO1	Subunit 1 of Cytochrome c oxidase (complex IV)
mtDNA	mitochondrial DNA
MTND4 and MTND5	Subunits 4 and 5 of NADH-ubiquinone oxidoreductase (complex I)
np(s)	nucleotide position(s)
OXPHOS	oxidative phosphorylation
PCR	polymerase chain reaction
PRIFW	putative recent import from the west
SCR	Southern Coastal Route
KYA	thousand (kilo-) years ago
MYA	million (mega-) years ago

Definitions of basic terms, used in current dissertation

Haplotype (= lineage)		mtDNA sequence with characteristic polymorphisms, 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		time estimate to MRCA;
MRCA		Most recent common ancestor
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

Recent decades have seen the blossoming of genomics. The completing of the human genome sequence and the progress of the HapMap project are its major manifestations – at least as far as our species is concerned. *Inter alia,* this era has proven extremely resourceful in terms of reconstructing the past of us – *Homo sapiens*. Here, haploid genomics has made the most significant impact into an essentially inter- and multidisciplinary field of research, often taken together in an evergreen question "Where do we come from, who we are, and where are we going?"

The literature overview of my thesis is built up from three parts. Firstly, I discuss recent developments regarding some of the key properties of mtDNA for phylogenetic studies. Secondly, I go into discussing the peopling of Asia, where I concentrate on the phase immediately following the initial out-of-Africa exodus *via* the Southern Coastal Route. Lastly I attempt to give a brief overview of Indian palaeontology and archaeology, to fly over South Asia from the linguistic point of view and to summarise, in a few words, essentials of the "pre-DNA" era of studies on genetic variation in India. Some of the questions and disagreements surfacing from the literature overview are subject to enquiry in the current study.

2. LITERATURE OVERVIEW

2.1. Some recent developments regarding the properties of mtDNA in respect to phylogenetic studies

The human mitochondrial genome (mtDNA) is a tiny (16.6 kbp) circular double stranded DNA molecule that contains blueprints for just 13 proteins, 22 transfer RNAs and two ribosomal RNAs (Figure 1) (Anderson et al. 1981; Andrews et al. 1999). The 1.1 kbp long segment called the control region (CR) is the largest region of the mtDNA which does not contain coding information. Otherwise, mtDNA is virtually devoid of noncoding DNA. The CR contains the control elements for replication and transcription of mtDNA and three hypervariable segments one to three– HVS I–III (nps: 16024–16365, 73–340 and 438–574, respectively) (Greenberg et al. 1983; Wilson et al. 1993; Lutz et al. 1998) which, as their names indicate, are in average more polymorphic than the rest of the genome.

Several features of the mtDNA differentiate it from the nuclear genome and make it a powerful tool for phylogenetic studies. MtDNA is inherited along the maternal line of decent (Giles et al. 1980) and does not recombine (Olivo et al. 1983; Merriwether et al. 1991, but see chapter 2.1.1). This enables reconstruction of a true phylogeny as a genealogy of individuals (carriers of distinct mtDNA haplotypes) – as opposed to nuclear marker systems where variation is expressed largely in the frequencies of different alleles. Reconstruction of the phylogeny, application of the coalescent theory to the phylogeny and linking this information to geography, archaeology, palaeontology etc. form the methodological basis for phylogeographic approach (Avise 2000), which is most widely used to study the past of humans as written in their mtDNA sequence variation.

Another feature of the mtDNA arising from its maternal inheritance and lack of recombination is its four-fold smaller effective population size (N_e) as compared to the nuclear loci. This makes mtDNA more susceptible to the effects of random genetic drift, which in turn allows us to "see" past founder effects and/or bottlenecks that might be undetectable at the level of nuclear loci. At its extremes, though, drift can erase signals of past demographic events by preventing us to "see" *through* a founder effect or a bottleneck – the two manifestations of random genetic drift.

The human mitochondrial genome evolves more than ten times faster than the nuclear genome (Brown et al. 1979; Ingman and Gyllensten 2001). This is another key characteristic of the mtDNA because it literally translates into more available information *per* nucleotide that can be studied to reconstruct the shallow history of our species. On the other hand, one must be aware of the other side of the coin, namely saturation – multiple mutations in the same position that would obscure the genealogy.

The excess of synonymous over non-synonymous substitutions in all mtDNA protein-coding genes shows that the genes are functionally active and that negative (purifying) selection eliminates non-synonymous substitutions as most of them are deleterious. Another question is if, and how, selection has shaped the topology of the phylogenetic tree. I shall return to this discussion in chapter 2.1.2.

These special features of the mtDNA have been discussed in depth also in textbooks (Jobling et al. 2004) and other PhD theses also defended before this council (Kivisild 2000; Tambets 2004) and will hence be not subject to an extended overview here. However, a few recent studies have questioned some of these key perquisite properties of the mtDNA for phylogenetic studies. Therefore, in the following two sections I shall touch upon two issues that have recently stirred up discussion: i) paternal inheritance and recombination of mtDNA, and ii) the role of selection in shaping the extant human mtDNA pool.

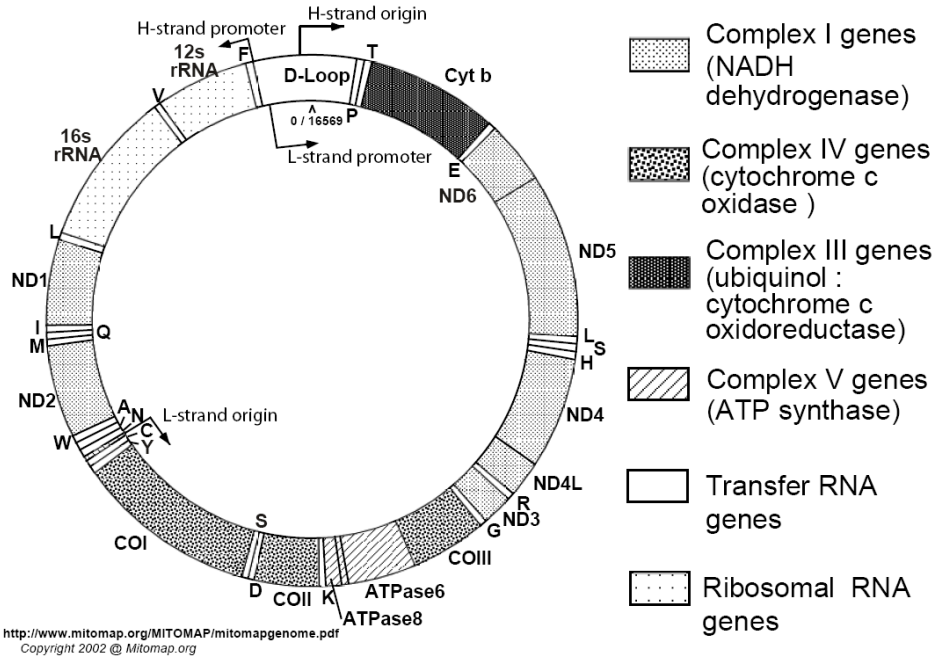

Figure 1. Human mtDNA genomic map (modified from www.mitomap.org). Genes transcribed from the H-strand and the L-strand are shown outside or inside the circle, respectively. Promoters for transcription and replication origins are indicated with arrows.

2.1.1. Mitochondrial Steve – a gloomy story of a substitute player who somehow never really got on the field thus far

Exclusively maternal inheritance and the lack of recombination of mtDNA are the cornerstones of mtDNA phylogenetic studies permitting the reconstruction of individual genealogies and their evolution through time. Paternal sperm-borne mitochondria of most mammalian species do enter the ooplasm at fertilization but are specifically targeted for degradation by the resident ubiquitin system. Prohibitin, the major protein of the inner mitochondrial membrane, appears to be ubiquitinated in the sperm mitochondria (Sutovsky et al. 2004). Paternal mtDNA leakage into the next generation is further hindered by the huge copy number difference between the ootcyte (100,000) and the sperm (50–1,200) and the bottleneck of mitochondria in the developing ootcyte (Koehler et al. 1991; Jenuth et al. 1996; Marchington et al. 1998; Poulton et al. 1998; Thorburn and Dahl 2001).

Still, from time to time studies claiming paternal inheritance in humans (Egger and Wilson 1983) pop up. Similar claims are also made for other mammals like, for example, sheep (Zhao et al. 2004). A recent report of paternal inheritance on the background of a severe mitochondrial metabolic disease has brought the matter under the spot light again (Schwartz and Vissing 2002). The presence of paternal mtDNA in the muscle cells of a 28 year old Danish patient has been discussed and potentially profound implications suggested (Bromham et al. 2002; Schwartz and Vissing 2003). Bromham and colleagues (2002) also coined the term Mitochondrial Steve as an analogy to Mitochondrial Eve. However, subsequent studies trying to spot mitochondrial disorder associated paternal transmission have so far failed (Filosto et al. 2003; Taylor et al. 2003; Schwartz and Vissing 2004). Moreover, there are no reports suggesting paternal mtDNA leakage in normal conditions.

Similarly, the issue of recombination among mtDNA genomes has been discussed now and then. There have been recent reports showing recombination in an i*n vitro* cell culture system (D'Aurelio et al. 2004) and following paternal mtDNA transmission in somatic tissue of the same Danish patient mentioned before (Kraytsberg et al. 2004). Bandelt and colleagues have screened a wealth of medical literature for signs of artificial mtDNA recombination and for an "evidence" for paternal transmission (Bandelt et al. 2004; Bandelt et al. 2005). They conclude that due to the lack of will/know-how to test mtDNA sequences data phylogenetically, claims for recombination and/or paternal transmission can in most cases be reduced to simple sample mix-ups or PCR contamination. Also the study of Kraytsberg and colleagues (2004) received criticism for its methodology (Bandelt et al. 2005). Nevertheless, a very recent report of recombination in human skeletal muscle cells in ten individuals with multiple mtDNA heteroplasmy (Zsurka et al. 2005), suggests that, given the opportunity,

in somatic tissue recombination may be a reality. Criticism on methodological grounds will be published shortly, though (Bandelt, personal communication)

A number of events are needed in tandem for a recombinant to enter the gene pool: paternal penetrance, the opportunity for recombination to take place, the passing on of the recombinant into the next generation and its spread in the population. While, as we saw, there is some evidence for the first two events, the results of the latter two would be testable and should be tested on the gene pool level. There have been indeed claims of interpreting the extant mtDNA variation partly as a result of past recombination events. A putative recombinant control region haplotype was reported from one pacific island population (Hagelberg et al. 1999). This claim was shortly withdrawn as the authors acknowledged mtDNA sequence alignment errors *in silico* (Hagelberg et al. 2000). Further, the decline of LD with the increase of distance between sites was reported as testifying for recombination in action (Awadalla et al. 1999). Similar interpretation arose from a study where largely the same authors claimed that the frequency of parallel mutations at the same nucleotide positions is much higher than expected on the basis of single rate of synonymous mutations (Eyre-Walker et al. 1999). These studies have however received strong criticism both on methodological (unsuitable LD statistics) and on data-quality grounds (errors in genotyping or data handling) (Macaulay et al. 1999a; Jorde and Bamshad 2000; Kivisild and Villems 2000; Kumar et al. 2000). These matters aside, recombination is not the only explanation for decay of LD. It can also be caused by mutational hotspots. However, a reanalysis of the data suggested that simple chance was behind the observed LD decay (Innan and Nordborg 2002). Despite the criticism on the evidence for recombination, discussion on its (hypothetical) implications continue to be published (Hey 2000; Hagelberg 2003). Others discuss methods that can and cannot in principle detect recombination (Wiuf 2001; McVean et al. 2002).

A number of additional studies have looked for signs of recombination potentially surfacing from the ever-growing body of mtDNA full sequences and seen none (Ingman et al. 2000; Jorde and Bamshad 2000; Elson et al. 2001; Piganeau and Eyre-Walker 2004). Therefore, we can conclude that although there is some evidence for isolated occasions of parental mtDNA leakage and, possibly, mtDNA recombination in somatic cells, there is no evidence suggesting that recombination has shaped the evolution of mtDNA in humans. First and foremost, there is so far no evidence for maternal transmission of paternally inherited mtDNA down to the next generation and into the gene pool. Hence, even if we were to assume that leakage of paternal mtDNA into the maternal decent is in principle possible, it would be an extremely rare event, whilst the fate of such a rare haplotype is very likely in its elimination by random genetic drift.

2.1.2. Positive selection towards heat production through less-efficient OXPHOS in the freezing North? Problematic landing for a nice idea

The large number of mtDNA full sequences made available during recent years have enabled discussion of the effects of selection on human mtDNA evolution (Ingman and Gyllensten 2001; Mishmar et al. 2003; Moilanen et al. 2003; Moilanen and Majamaa 2003; Elson et al. 2004; Ruiz-Pesini et al. 2004; Kivisild et al. 2005). The ratio of synonymous to non-synonymous substitutions (at different depths of the human mtDNA phylogeny) indicates that purifying (negative) selection has in fact shaped human mtDNA evolution – an observation which is only essential to assume. More interestingly, it appears that amino acid replacements in the periphery of the phylogenetic tree are more deleterious than those in the central parts (Moilanen and Majamaa 2003), yet again suggesting negative selection in action. This, in turn, may affect "handling" the molecular clock, applied over the complete genomic sequence of mtDNA (Ho et al. 2005; Kivisild et al. 2005). To overcome this problem Kivisild and colleagues (2005) provided a new calibration method for the mutation rate of synonymous transitions to estimate the coalescent times of mtDNA haplogroups. As calibrated over the observed transversion rate at synonymous and rRNA positions between human and chimpanzee the average mutation rate over all human mtDNA genes yielded one synonymous substitution per 6,764 years (Kivisild et al. 2005).

Most studies on mtDNA phylogeography have considered the regional distribution of mtDNA haplogroups as a result of genetic drift and random accumulation of new mutations. However, some recent studies have challenged this view pointing to adaptive (positive) selection as an important player (Mishmar et al. 2003; Ruiz-Pesini et al. 2004). Comparing the most frequent haplogroups in Siberian (cold), European (temperate) and African (tropical) populations, the authors reported that there was an excess of non-synonymous substitutions at the bases of these haplogroups as one moves towards the colder climate, indicating for adaptation rather than drift (Ruiz-Pesini et al. 2004). The authors argued that because the non-synonymous mutations would be (mildly) deleterious for the OXPHOS system, the balance between producing ATP and heat in the mitochondria would become inclined towards heat production at the expense of ATP. This, they continued, would be advantageous in the freezing north – a mutation selected for. And indeed, higher metabolic rate among the circumpolar populations has been observed (Leonard et al. 2002) and was discussed as supporting evidence (Mishmar et al. 2003; Ruiz-Pesini et al. 2004). However, the most frequent haplogroups in Siberia are much (by tens of thousands of years) younger than the L clade, taken by them to represent the tropical Africa. Therefore the excess of non-synonymous substitutions in the former may have an explanation which does not require involvement of positive

selection. Indeed, comparison of the nature of substitutions in haplogroups of similar age/diversity of different climatic regions does not reveal reliable differences (Table 4 of Kivisild et al. 2005).

Adaptation to climate, as a possible significant force in mtDNA evolution, has been questioned also by others (Moilanen et al. 2003; Elson et al. 2004). However, these studies have reported lineage-specific differences in selective constraints in different regions of the human mtDNA (Moilanen et al. 2003). For example, maximum-likelihood sliding window analysis indicated that regions with the highest diversity in the mtDNA differ between haplogroup clusters. But, perhaps the most notable result of Moilanen and colleagues (2003) was significant under-representation of non-synonymous substitutions in mtDNA region between nps 12478–13611 in hg J background. This region codes for amino acids 48–425 of MTND5. This hydrophobic polypeptide belongs to the membrane-spanning part of the complex I, and it is probably an important component of the proton translocation machinery (Mathiesen and Hagerhall 2002). The authors suggest that the respective amino acids might have been under stronger selective pressure in hg J than in others, resulting in the observed lack of mutations. Intriguingly, hg J is defined by three amino acid replacements, two of which are within the subunits of complex I and are therefore possible candidates for mutual interactions.

The intensity of selection varies also between different genes encoded by the mtDNA. For example MTCO1 and MTND4 have beet shown to harbour less non-synonymous substitutions than expected (consistent with negative selection) (Elson et al. 2004). In contrast, MTATP6 gene showed an excess of mutations that change the amino acid residue (at least in European and Asian populations), thus pointing to positive or balancing selection (Mishmar et al. 2003; Elson et al. 2004).

It has to be noted, though that while the relative excess of and shortage of synonymous substitutions have been interpreted as negative and positive selection, respectively, they could, in fact, also be interpreted as relaxation of selective constraints and relaxation of negative selection, respectively (Elson et al. 2004).

Another interesting phenomenon was reported by Kivisild and colleagues (2005). They found a significant excess of non-synonymous substitutions involving threonine and valine codons in the human mtDNA. Moreover, the ratio of these changes from and to threonine and valine differed significantly between populations. The authors considered differences in diet as one potential explanation. Threonine and valine are both essential amino acids that are not synthesised and must therefore be consumed with food. Both amino acids are plentiful in meat, fish, lentils, peanuts, and cottage cheese but rare in most grains. Could the change from hunting and gathering to agriculture as the main mode of subsistence drive the replacement of threonine? For example in haplogroup V sequences from Finland, where people continued to rely largely on hunting and fishing for subsistence even long after the first contacts with

farming, changes to threonine and valine are prevalent while for H1 sequences sampled from Europe and the Near East the opposite is true (Kivisild et al. 2005).

In conclusion, there are numerous hints for non-random processes in the evolution of the human mtDNA, and the subject warrants further careful investigations and interpretations. For example, the use of an average clock for all sites over all genes of the mtDNA might be only the first approximation – to be refined in the course of the accumulation of more experimental data. However, the departures from neutrality do not undermine the use of the phylogenetic approach for analysing mtDNA sequence data as such, not to add that the reconstruction of the basal branches of the tree is robust and the excess of non-synonymous substitutions affects mainly the tips of the tree.

2.2. The pioneer settlement of modern humans in Asia

A note:
With minor adaptations this chapter of the Literature Overview has been submitted to publication:

Metspalu M, Kivisild T, Bandelt H-J, Richards M, Villems R (in press) The pioneer settlement of modern humans in Asia. In: Bandelt H-J, Macaulay V, Richards M (eds) Human mitochondrial DNA and the evolution of Homo sapiens. Springer-Verlag, Heidelberg

2.2.1. Introduction

Different hypotheses, routes, and the timing of the out-of-Africa migration are not the focus of this chapter. However, in order to dig more deeply into discussions about pioneer settlement of Asia, it is necessary to emphasize here that many recent genetic, archaeological, and anthropological studies have started to favour the Southern Coastal Route (SCR) concept as the main mechanism of the primary settlement of Asia (Lahr and Foley 1994; Quintana-Murci et al. 1999; Stringer 2000; Kivisild et al. 2003; Kivisild et al. 2004); see also Oppenheimer (2003).

The coastal habitat as the medium for humans to penetrate from East Africa to Asia and Australasia was perhaps first envisaged by the evolutionary geographer Carl Sauer, who considered the populations taking this route as being adapted to the ecological niche of sea shore (Sauer 1962). After reaching southwest Asia, modern humans had a choice of two potential routes by which to colonize the rest of Asia. These two were separated by the world's mightiest mountain system – the Himalayas. The pioneer settlers could continue taking

the SCR or they could change their habitat and turn instead to the north, passing through Central Asia and southern Siberia (or via the route that later became known as the Silk Road). Here, one has to avoid confusion with the "Northern Route" of the Out-of-Africa exodus and use the term "Northern Asian Route".

In principle, the first pioneer population of modern humans could have spread in several directions simultaneously: for example following the coast towards the east and, at the same time, cutting into the Asian inland along the river valleys (Wells et al. 2001; Oppenheimer 2003). As we would argue below, a single coastal route towards the east appears to be sufficient to explain most, if not all, existing mtDNA genetic variation, not only in North and Southeast Asia, but also in Oceania. Before discussing extant mtDNA diversity in Asia, we shall briefly review the palaeoclimatological background and archaeological evidence for these events.

2.2.2. Palaeoclimatological context

The patterns of colonization of Asia by anatomically modern human (AMH) population(s) were undoubtedly highly dependent on the surrounding environment. Our knowledge of past climates is frozen in ice sheets: polar ice cores and (ocean) sediments have been intensively studied in order to reconstruct the climate of the past. These global changes are reflected in the more detailed regional palaeovegetation surveys based, for example, on ancient pollen analysis (reviewed in Adams and Faure; Adams et al. 1999; Ray and Adams 2001). As temperature change is a robust characteristic of the environment, we shall concentrate on that and will not go in details of, say, palaeovegetation.

After the Eemian interglacial, some 110–130 thousand years ago (KYA), the global climate cooled until the period of the lowest temperatures, during the Last Glacial Maximum (LGM) ~15–25 KYA. This process was not a steady one; instead, there were multiple oscillations of warmer and chillier periods. Intense cold and arid, but short-lived, Heinrich-type events characterized the otherwise gradually cooling phase between 110 and 70 KYA. This was followed by the Stage 4 Glacial Maximum (also known as the Early Wisconsin Glacial) extending to ~50 KYA with conditions rather similar to the LGM (Adams et al. 1999). Warmer but highly variable temperatures were characteristic of the period thereafter, extending until the onset of LGM.

As much as the changing temperatures, the peopling of Asia by anatomically modern humans was affected by the accompanying fluctuations in humidity. Lower temperatures mean generally less evaporation. The resulting global decrease in rainfall contributed to the extension of desert areas, e.g. in Central and south-western Asia. Even ~50 KYA, when a warmer and moister stage opened the green passage between the Arabian Sea and the Levant (the Zagros corridor), the deserts in Central Asia and northern Africa remained difficult

habitats for most of the creatures – including humans (reviewed in Oppenheimer 2003).

During the colder phases of climate, which were also much more arid, the global sea level was much lower than it is today. That was mainly because enormous quantities of water were trapped in the extended polar ice caps, with a corresponding decrease in the volume of the oceans as the water cooled. The fluctuation of the sea level made crossing water obstacles easier at some times and harder at others. For example, the distance between Australia and Timor (the widest strait that had to be crossed en route to Sahul, or the Greater Australian landmass comprising both Australia and New Guinea during the Ice Age) was shortest (at 170 km) between 65 and 70 KYA, when the global sea level was about 80 meters below its present level. During the subsequent warmer phases the strait lengthened, but did not exceed 220km even during the maximum high-stand (~50 KYA) when the sea stood only 40 meters below its current level (Chappell 2002). However, given the development of some form of water-craft, there seems no strong reason to regard the time of shortest distance as the sole window of opportunity for crossing (cf. Oppenheimer 2003).

2.2.3. Archaeological and palaeontological evidence of the peopling of Asia by AMH

Any fossil record is, inevitably, incomplete. Fossilization of skeletal remains is a rare event, depending on climate and probably many other factors, whereas their recovery depends on the intensity of archaeological investigation of a region. There is a particular issue when one considers the course of the likely coastal route out of Africa. If the beachcombing modern humans, being dependent on seashore environment, indeed began colonizing Eurasia via the SCR, then many of the potential archaeological sites are at present submerged under the sea. An 80m rise of the sea level (the difference between 70 KYA and today) altered the coastline considerably, shifting it in some places hundreds of kilometres inland and probably inundating the range of beachcombing AMH. Another important factor that needs to be considered is the tectonics of the continental shelves. Furthermore, the accuracy of fossil dating techniques is under constant dispute (see Chen and Zhang 1991; Klein 1999).

Despite these problems, fossils are and will probably continue to be the best evidence of AMH spread around the globe. Widely accepted datings for the earliest AMH skeletal remains outside Africa (excluding the brief "extension" of the African range into the Near East during the Eemian interglacial) reach approximately 45 KYA (Foley 1998 and refs. therein). Claims for considerably older dates in Asia, e.g. 67 KYA for Liujiang, China, have been heavily criticized (e.g. Etler 1996), although they continue to be made (Shen et al. 2002).

In Europe, the earliest AMH remains fall between 37 and 45 KYA (van Andel et al. 2003). Interestingly, AMH remains of similar antiquity have been found in Borneo and Australia, where the most ancient remains at Niah Cave and Lake Mungo, respectively, have recently been re-dated to >40 KYA (Barker et al. 2002; Bowler et al. 2003). The beginning of the human settlement at the Australian site was estimated to go back even to around 50 KYA (Bowler et al. 2003) or even ~62 KYA (Thorne et al. 1999) while similar dates have been proposed for other Australian sites, such as Devil's Lair, south-western Australia (Turney et al. 2001), and Deaf Adder Gorge, northern Australia (Roberts et al. 1990; Roberts et al. 1994). These dates are of great importance because they also set the time boundary for the peopling of Asia.

The fact that younger remains from Inner Mongolia, with probable antiquity of not much over 30 KYA, constitute the earliest widely accepted fossil findings of AMH in mainland East Asia (Chen and Zhang 1991; Etler 1996) highlights the weakness of negative arguments raised by the lack of fossil evidence. Rather similar is the situation in South Asia, where the oldest fossils of modern humans uncovered so far, from southern Sri Lanka, are dated to 28 and 33 KYA (Kennedy et al. 1987; Kennedy and Deraniyagala 1989)—the early dates of fossils in Australia imply that AMH had to be in South Asia at least 10 KY earlier than this.

Archaeological evidence has sometimes been interpreted as supporting modern human presence in South Asia over 60 KYA (references in Kumar and Reddy 2003). However, because the putative Middle Palaeolithic sites under consideration lack any human fossil evidence, it is not clear whether they can unambiguously be associated with modern humans (Joshi 1996). The spread of modern humans in the Middle East and Europe is generally coupled with the radiation of Upper Palaeolithic technology – which, however, did not reach Australia together with AMH. The same may well be true for South Asia, where the Upper Palaeolithic technology does not show up before 30 KYA (Chakrabarti 1999). Again, the introduction of Upper Palaeolithic technology to India postdates significantly the time frame when the carriers of the Middle Palaeolithic tools would have been walking on its shores to reach Australia. And, more importantly, when the Upper Palaeolithic reached India, it was contemporaneous with the pre-existing Middle Palaeolithic there for at least 10 thousand years.

The transition from Middle to Upper Palaeolithic in the southern Near East occurred ~50 KYA (Gilead 1991). In Europe, it emerges almost simultaneously in both central Europe and in northern Spain before expanding through the continent (~47 KYA: van Andel et al. 2003). The earliest Upper Palaeolithic technology is of nearly similar antiquity to the east, in the Zagros mountains (Olszewski and Dibble 1994) but slightly younger in the Caucasus region (30–32 KYA: Bar-Yosef 2001), where Neanderthals survived until ~32 KYA. A similarly early transition (39–43 KYA) has been suggested for the Altai Mountain and Lake Baikal regions of southern Siberia (Dolukhanov et al. 2002;

Vasil'ev et al. 2002), although these Early Upper Palaeolithic cultures share many features in common with the preceding Middle Palaeolithic Mousterian culture (Kuzmin and Keates 2004). Moreover, the archaeological record alone, with a lack of human skeletal remains, is inconclusive regarding whether or not the initial Middle to Upper Palaeolithic transition in Siberia was coupled with the influx of an AMH population from the west. Upper Palaeolithic artefacts from 18 KYA have been found in association with skeletal remains that bear similar morphology to contemporary AMH teeth from Europe (Scott and Turner 1997).

The Middle Palaeolithic settlement of AMH in Sahul, together with the radiation of Upper Palaeolithic technology from the Middle East and its early arrival in southern Siberia, have often been interpreted as supporting the existence of two different migration routes from Africa towards East Asia: an earlier one following the southern route along the coast of Asia towards Australia, carrying Middle Palaeolithic technology, followed by a migration associated with the Upper Palaeolithic via the northern route through the Levant and further along the Northern Asian Route through Central Asia and southern Siberia (Lahr and Foley 1994; Jobling and Tyler-Smith 2003). This twin-dispersal model offers different testable predictions about genetic patterns to be observed. If the two-route scenario (or "pincer model") would indeed explain the source of modern humans in Asia, then one should be able to find unique Northern and Central Asian-specific lineages that cannot be derived from South and Southeast Asian variation, and *vice versa*. On the other hand, if the single southern or northern route scenario holds, then it should be possible to derive all northern variants from the gene pool of the south, and *vice versa.*

2.2.4. How to reconstruct "pioneer settlement" from extant MtDNA diversity

An obvious starting point for deducing the patterns of the pioneer human settlement from the extant mtDNA diversity is to identify regionally autochthonous haplogroups and calculate their coalescence ages. The average over the oldest of these would indicate the lower bounds for the start of the colonization. A founder type is identified as an ancestral node which is present (or may have been lost but is then phylogenetically reconstructed) both in the source and the destination area (Richards et al. 2000). Ideally, the coalescence time of the founder type in the destination area would suggest the time of its arrival (Stoneking et al. 1990; Torroni et al. 1993a; Torroni et al. 1993b; Sykes et al. 1995; Forster et al. 1996; Richards et al. 2000). However, let's look at two non-ideal cases. Firstly, if the founder population is small and does not disperse/expand upon arrival, the coalescence times of the founder types may underestimate their entrance time. This is because the most recent common

ancestors for the future generations may successfully be replaced by younger ones, which is evinced by the phylogenetic reconstructions of branches defined by multiple mutations. Most likely these mutations occurred one by one in the evolutionary sequence, but in the extant populations it is not clear which mutation occurred before and which after the founding event of that particular region of interest. For evaluation and comparison, one can also draw the tree and calculate the age of the descending haplogroups in the supposed source area. Secondly, in more massive migrations, a considerable amount of variation (within a haplogroup) may already be present among migrants and in that case their extant diversity (per haplogroup) is a sum of different periods of their demographic history. Back-migration(s) to the source area, and the ability to detect founder types through adequate sampling, are the other two major challenges for the founder analysis approach (Richards et al. 2000).

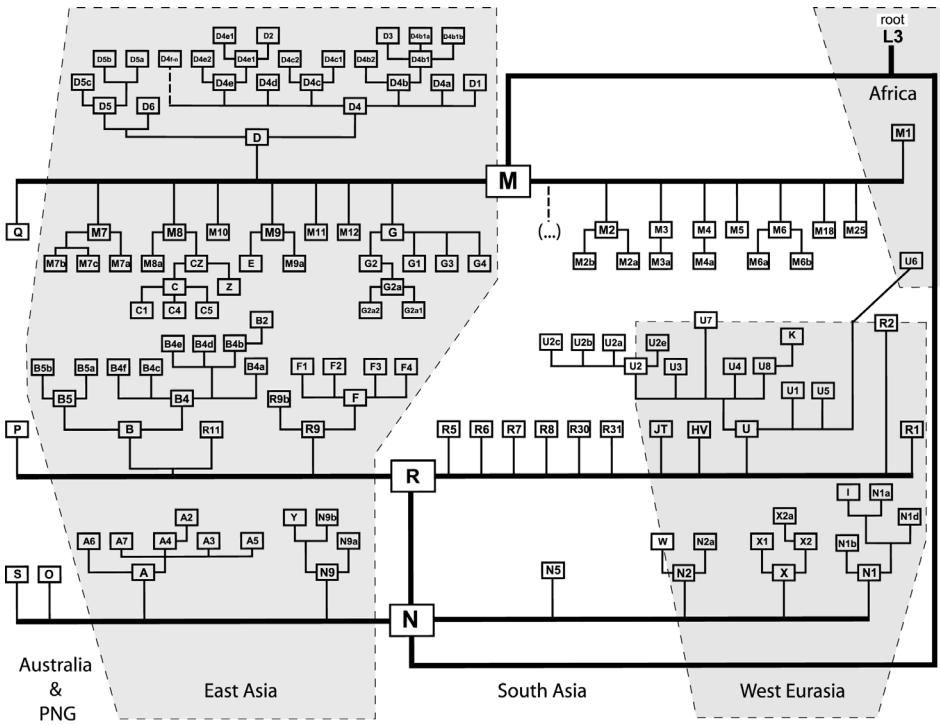

Figure 2. Phylogenetic tree relating the Asian mtDNA haplogroups. See www.evolutsioon.ut.ee for further information, particularly regarding the nomenclature of the haplogroups.

The outlined strategy requires good data coverage of the regions under investigation. Good in this context means both the "depth" – phylogenetic resolution – and "width" – geographic scope – of the data sets. Phylogenetic depth can be improved by searching for more markers until the bounds are met for the specific locus – a step that is already achieved for human mitochondrial DNA, but still lies ahead for the Y chromosome, for example. In practice, however, when surveying mtDNA diversity in different regions, one is quite often limited by the depth of the available data sets, since these often still only consist of HVS-I sequences, which are still popular in molecular anthropology as "the high mutation rate of this segment ensures a sufficient number of polymorphic sites for population genetic analyses" (Pakendorf and Stoneking 2005). It is exactly the recurrent nature of most mutations in this segment that inevitably destroys the more ancient signals one is usually interested in. Such data sets are thus of a limited value in the absence of coding-region information, and one should be careful with making inferences based on insufficient information.

2.2.5. The peopling of Asia as seen through the lens of mtDNA diversity

Complete and partial mtDNA coding region sequences have been used to map the backbone and determine the fine-structure of the mtDNA lineages present in Asia (Kivisild et al. 2002; Yao et al. 2002; Kong et al. 2003; Metspalu et al. 2004; Palanichamy et al. 2004; Quintana-Murci et al. 2004). The recent analysis of complete mtDNA sequences from 672 Japanese individuals has provided a significant refinement of the East Asian mtDNA phylogeny (Tanaka et al. 2004). Combining these and other published data, Figure 2 summarizes the Asian mtDNA tree topology. With many papers that refine the phylogeny being published almost simultaneously, it is hardly surprising that relevant literature may sometimes be missed and some confusion regarding to the naming of the haplogroups and their branching order arises. The ongoing flow of complete mtDNA sequences from newly emerging basal branches within the Asian-Oceanian mtDNA phylogeny could not find place in this single diagram. A more detailed version of Figure 2, along with commentaries that aim to overcome some of these difficulties, can be obtained at www.evolutsioon.ut.ee/. The macrohaplogroups M and N effectively cover the whole mtDNA pool in Asia. The start of their dispersal has been dated to approximately 60–65 KYA (Maca-Meyer et al. 2001; Mishmar et al. 2003; Palanichamy et al. 2004). Macrohaplogroup M is slightly more frequent than N in Siberia, northern China, Japan and South Asia, while in Southeast Asia it is the other way around. M is practically absent from Southwest Asia where subhaplogroups branching from the N (including R) trunk dominate the mtDNA landscape. The N and R sub-branches in West and East Eurasia do not overlap, thus forming two distinct mtDNA "domains". With approximately similar shares, these two make up most of the mtDNA pool of Central Asia (Comas et al. 2004).

While the mtDNA makeup of the Americas represents an offshoot of the East Asian domain (Torroni et al. 1993a; Torroni et al. 1993b; Torroni et al. 1994; Forster et al. 1996), Sahul (Australia/New Guinea) largely constitutes yet another autochthonous one. Stemming from the trunks R and M, haplogroups P and Q, respectively, cover more than half of the extant mtDNA pool sampled in Papua New Guinea (Forster et al. 2001). From the published full sequences (Ingman and Gyllensten 2003), additional Sahul-specific basal branches of M and N – namely, S and O, which were baptized in (Palanichamy et al. 2004) – are confirmed. Thus, both basal trunks of the Eurasian mtDNA tree show deep-rooting Sahul-specific branches.

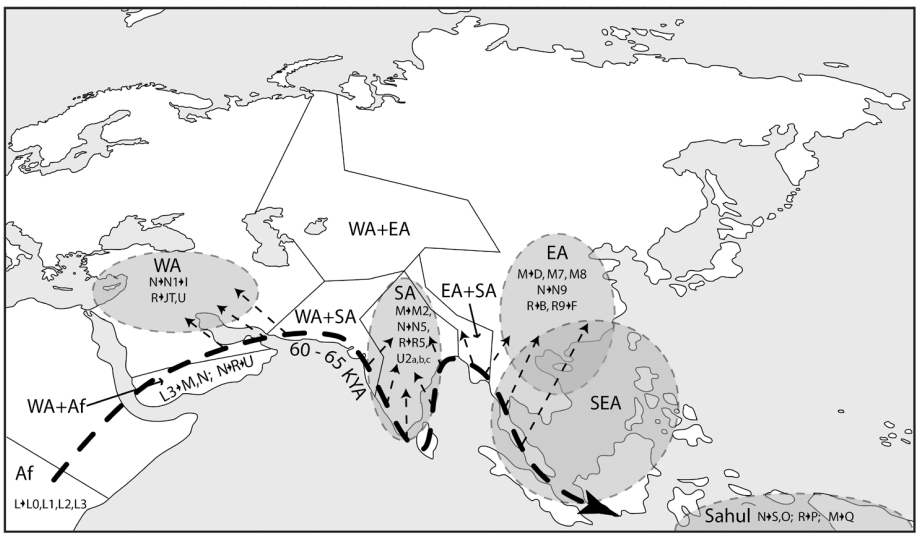

Figure 3. Map of Eurasia depicting the possible scenario of the pioneer settlement of modern humans in Asia. The heavy dashed arrow pictures the initial southern (coastal) route of the out-of-Africa event which had taken place by around 60–65 KYA (Af = African specific mtDNA variants). During this opening stage, the earliest offshoots of haplogroups M and N were rapidly segregated into West (WA: e.g. JT, U), South (SA: e.g. M2, N5, R5, U2a, U2b, U2c), East Asian (EA: e.g. D, M7, M8, N9, R9, B) and further into the Australasia-specific (Sahul: S, O, P, Q) variants which later became the inocula for the autochthonous mtDNA diversification in the respective regions (light dashed arrows and ellipses, Sahul not fully shown). The complete mtDNA sequence data suggests a number of autochthonous Southeast Asian (SEA) M and N lineages that are absent from northern East Asia. During later stages of the colonization of Eurasia, modern humans moved further inland (not shown). Admixture between these basic domains of human settlement in Asia has been surprisingly limited on the maternal side ever since. Approximate boundary admixture zones (over 20% of admixture) between the three domains are shown by crude solid lines together with an indication to the mixed domains. Note that the whole of Central Asia appears as the biggest admixture zone where the mtDNA pools of West and East Asia, and to a very much lower extent South Asia, intermix.

South Asia, with its own specific branches of M and N, represents the third mtDNA domain in Asia (Figure 2). Haplogroups M2, R5 and subgroups U2a, U2b, and U2c of U2 (Kivisild et al. 1999a; Kivisild et al. 2003; Quintana-Murci et al. 2004), which make up more than 15% of the South Asian mtDNAs, each show coalescence times of over 50 KYA (Metspalu et al. 2004). These haplogroups form a set of the most ancient Indian-specific haplogroups identified so far. A number of novel Indian-specific basal N and R lineages (N5, R7, R8, R30 and R31) were recently identified from complete sequences (Palanichamy et al. 2004). The phylogeography of these in South Asia needs further attention, but, significantly, their autochthonous presence in India clearly demonstrates that all the basal trunks – M, N and R – have diversified in situ. The coding-region based downstream classification of haplogroup M lineages in South Asia still needs further effort, as most are still grouped into the paragroup M*. However, it is already clear that they differ from the identified M subhaplogroups of East Asia (Kivisild et al. 1999b; Metspalu et al. 2004).

Overall, then, the South Asian mtDNA pool consists of autochthonous branches of the global mtDNA tree that stem directly from each of the basal trunks M, N and R (Figure 2). Note that the only major Indian-specific lineages not stemming directly from the trunk are the Indian subhaplogroups of U2, which may have a sister group U2e in West Eurasia (although we note that this putative sister relationship hinges upon a single transition, at nucleotide position 16051 in the control region, which may not have been a unique event at the base of haplogroup U). The divergence time of these U2 daughters reaches ~50 KY (Kivisild et al. 1999a). Meanwhile, haplogroups R2, U7 and W represent an intriguing link between the West and South Asian mtDNA pools. Their spread and coalescence times suggest pre-LGM gene flows in the area spanning from western India and Pakistan up to the Near East. As judged from the coalescence times of the region-specific subclades of these haplogroups, this genetic continuum was apparently interrupted by the expanding deserts in eastern West Asia during the LGM (Metspalu et al. 2004).

Like Sahul and South Asia, the East Asian mtDNA pool is made up of autochthonous offshoots of M and N, most of which show coalescence times exceeding 50 KYA (Kivisild et al. 2002; Yao et al. 2002). While in South Asia we see a number of basal haplogroups branching from the trunk R, in East Asia, only a few haplogroups, B (plus R11) and R9 (including F), spring out from the founder haplotype of haplogroup R (Figure 2). The putative monophyly of a supergroup R11'B is based solely on a transition at unstable nucleotide positions 16189 and 16519, which could very well have happened in parallel. Furthermore, no more than two East Asian haplogroups, A and N9 (including Y), trace back to MRCAs in the N trunk. As in South Asia, the richest trunk in East Asia in terms of haplogroups stemming from it is M. Complete mtDNA sequencing has indicated that Southeast Asia also harbours some autochthonous M, N and R lineages apparently not found further north, in East Asia (Macaulay et al. 2005; Merriwether et al. 2005).

As we see, all of the three mtDNA domains along the SCR (South Asia, East Asia and Sahul) harbour haplogroups that stem directly from the M, N and R trunks (Figure 2), are primarily spread only within a single domain, and demonstrate coalescence times comparable to the initial expansion of M and N. They are frequent enough to rule out the possibility of major gene flow between the domains (since these haplogroups are not shared between the domains).

A plausible model for the initial peopling of Asia, one might think, would be a series of nested daughter colonizations, where regions were peopled one by one, with a time lag. In such a case, one would observe East Asian haplogroups to be derived from South Asian haplogroups, and Sahul daughter clades from East Asian haplogroups. This is the case with the later colonization of the Americas, but evidently not along the SCR, where deep-rooting autochthonous branches of the mtDNA tree are present throughout. It suggests that the initial colonization of Asia was not a gradual process, but rather a swift one, spreading the same founder types along the shores of the Indian Ocean as far as Sahul (Figure 3). The fact that all of the domains show autochthonous basal R haplogroups – e.g. JT and U in West Asia, R5 (and many more) in South Asia, R9 and B in East Asia and P in Sahul – suggests that, in addition to the differentiation of M and N, the divergence of R also occurred at the start of the AMH expansion along the SCR. In addition, autochthonous lineages of haplogroup U2 trace back to ~50 KY both in West and South Asia. This time frame is comparable to the coalescence dates of M and N, possibly placing the spread of haplogroup U within the initial wave of the peopling of Asia. This would, however, demand attributing the absence of haplogroup U east of India to loss through genetic drift in the probably small scout population(s). Alternatively, a later, perhaps Upper Palaeolithic diffusion (starting from ~30 KYA: Chakrabarti 1999) from the West, ultimately from the Middle East (the most probable source of haplogroup U), might have introduced the U2 lineages into India (Kivisild et al. 1999a; Kivisild et al. 2000).

Here, it is worth briefly touching upon the "parallel world" of Y-chromosome variation. As with mtDNA, South Asia congregates a diverse set of the basal Eurasian Y-chromosome founder lineages – C, F and K. Further, a number of deep-rooting subclades of F are only distributed in South Asia (Kivisild et al. 2003). The package of Y-chromosome founder lineages in West Eurasia is reduced to F and K. These observations support the idea that the out-of-Africa migration first reached South and Southwest Asia and from there dispersed both east and west, consistent with the single SCR scenario. In addition more recent migration(s) from Africa, probably following the route over the Sinai, have enriched the West Eurasian gene pool with additional paternal lineages of haplogroup E (Underhill et al. 2001; Cruciani et al. 2004; Luis et al. 2004; Semino et al. 2004). A study of 19 Y-chromosome biallelic markers amongst East Asian populations revealed that the southern populations are more diverse than those from the north, the latter essentially representing a subset of the variation present among the former. This is consistent with the SCR scenario,

suggesting that mainland Southeast Asia was the starting point for the peopling of East Asia some 60 KYA (Su et al. 1999).

2.2.6. A route through Northern Asia?

Extant mtDNA variation in Asia suggests that the Southern Route was the major, if not the only, course used by the initial colonizers of Asia. But did some of the initial settlers of Asia populate first Central Asia and reach East Asia via southern Siberia (Maca-Meyer et al. 2001; Wells et al. 2001; Oppenheimer 2003; Tanaka et al. 2004)? The distinction between northern and southern East Asia seen in other biological traits (for example the spread of "sinodontic" and "sundadontic" teeth: (Scott and Turner 1997) is also evident in the mtDNA variation. The two East Asian haplogroups of the mitochondrial R trunk (B and R9) are predominantly of Southeast Asian provenance, while those with the MRCA in its ancestral N trunk (A, N9) are more frequent in northern East Asia. Unaware of the newly identified Indian-specific branch of N, N5 (Palanichamy et al. 2004), Tanaka et al. (2004) have interpreted the apparent lack of basal N lineages in India, and their presence in northern East Asia, as a strong argument for the Northern Asian Route for the peopling of Asia by a supposed "N population". However, when the gap in basal N lineages in South Asia is erased, the higher frequencies of some basal N lineages along the Northern Asian Route alone are not enough to corroborate its role in the process of peopling the continent. In addition, Tanaka et al. (2004) overlooked the phylogenetic relatedness of N and R and the abundance of basal R lineages in India.

Similarly, taken as a whole, the M trunk is more frequent in northern than in southern East Asia, but at the subhaplogroup level the picture is more complicated. For example, M7 and E, are largely specific to mainland and island Southeast Asia (Ballinger et al. 1992) while others like G, M8 (including C and Z), and the most frequent M subhaplogroup D, are much more frequent in northern East Asia. Haplogroups C and D are co-dominant in southern Siberia (Derenko et al. 2003) while C and G are more frequent in north-eastern Siberia (summarized in Tanaka et al. 2004). However, the mtDNA pools of northern and southern East Asia overlap, and the haplogroups that are most frequent among the Siberian populations also amount to one-quarter of the Southeast Asian mtDNA pool. In turn, the south-eastern haplogroups (excluding E, which is absent) take a notable share of the East Asian-specific mtDNAs in Central Asia (~22%) and southern Siberia (~13%).

Is this pattern a result of two separate initial migration routes, carrying different founder types, or a unilateral diffusion, followed by genetic drift? Central Asian- and southern Siberian-specific basal branches of M and N would be the "smoking gun" for the Northern Asian Route. Such branches have, however, not yet been found (Derenko et al. 2003; Comas et al. 2004). In the

absence of direct evidence, a more detailed analysis of the phylogeography of the mtDNA haplogroups that make up the East Asian share of the mtDNA pools of Central Asia and southern Siberia ought to help corroborate or rule out the existence of the Northern Asian Route. More specifically, these pools should be checked for haplogroups ancestral to at least some mtDNA lineages in East Asia.

Before we go any further, we should remind ourselves that the extant mtDNA pools of southern Siberia and, especially, Central Asia are mixtures of East and West Eurasian mtDNA domains (Derenko et al. 2003; Bermisheva et al. 2004; Comas et al. 2004). As one would expect, the share of the western haplogroups diminishes as we move eastward (>40% in Central Asia, <20% in southern Siberia and ~1% in East Asia). This pattern has been shaped by admixture along the Steppe Belt long after the initial peopling of Asia and, thus, lies outside the scope of this chapter. In the following discussion we shall, therefore, concentrate on the eastern Eurasian-specific share of the maternal lineages in Central Asia and southern Siberia.

Overall, the most frequent subhaplogroup of M in eastern Eurasia is D, which further branches into D4, D5 and D6 (Figure 2). We see decreasing representation of these as we move westwards from East Asia. Only haplogroup D4 has dispersed into Central Asia (inferred from Comas et al. 2004) whereas the frequency of D5 in southern Siberia (1.5%: Derenko et al. 2003) is fivefold lower than that in China (5–10%: Yao et al. 2002). This pattern itself is best explained by an East Asian origin of haplogroup D, a pattern that recurs for other haplogroups present in both East and Central Asia. Here we illustrate this reasoning by taking a more detailed look at several examples, starting with haplogroup D4.

Haplogroup D4 accounts for a third of the East Asian mtDNA lineages in Central Asia and a quarter of those in southern Siberia. Similarly, the frequency of D4 stays ~25% in northern China while it drops to ~10% in the south of the country (Yao et al. 2002) and even more as one travels further south (island Southeast Asia) or west (Indo-China). This pattern might seem consistent with the spread of D4 from Central Asia. However, a more detailed phylogeographic analysis questions that view.

One out of the myriad of D4 subhaplogroups found in East Asia (Figure 2), D4c accounts for ~40% of D4 in Central Asia (re-inferred from the HVS-I data of Comas et al. 2004). By comparison, for example, in southern Siberia the share of D4c in D4 is ten times smaller. In Japan, where D4c makes up ~13% of D4 (inferred from Maruyama et al. 2003), we see additional branches of the haplogroup, such as D4c1a (Tanaka et al. 2004), which seems to be absent in Central Asia (as judged from the absence of HVS-II motif 194–207). Other Central Asian D4 (HVS-I) haplotypes have close or exact matches in China, southern Siberia and tribal populations of eastern India. Hence, the palette of D4 subhaplogroups in Central Asia appears poorer than that in more eastern regions. This is consistent with an eastern origin of the haplogroup.

The situation is similar in several other haplogroups. Like D4c, G2a1a, (G2a in the original publication: see www.evolutsioon.ut.ee for clarifications) also shows high frequency in Central Asia, and is virtually the only variant of G found there (Comas et al. 2004). In southern Siberia, G2a1a constitutes a third of G while its sister clade G2a2 (3%), mother clade G2a (24%), and two other G subclades G1a1 (36%), G3 (6%) make up the other two thirds (Derenko et al. 2003) (see Figure 2). G2a1a is, along with its sister-clade G2a1b, also present in Japan. Furthermore, there is a subgroup of G2a1a in Japan, defined by transitions at np 16194 and 16195 (Maruyama et al. 2003; Tanaka et al. 2004). As we see, only one subhaplogroup, one bough of the G phylogeny, predominantly represents the dispersal of this haplogroup into Central Asia (Figure 2). This shows clearly that the phylogeography of neither haplogroup D nor G can be interpreted as supporting the origin of major East Asian-specific haplogroups in Central Asia.

This conclusion is further supported by the phylogeography of haplogroup M8 (Figure 2). This is the most frequent haplogroup in southern Siberia, accounting for around one half of the East Eurasian mtDNA lineages (Derenko et al. 2003). It is also relatively more frequent in Central Asia than in the east of the continent. However, at the basal level, we see again that in East Asia the presence of the various subhaplogroups of M8 (Figure 2) is more balanced, as C, Z and M8a are spread in comparable frequencies (e.g. among Han Chinese: Yao et al. 2002); whilst in southern Siberia, and particularly in Central Asia, subgroups C and Z are predominant. Therefore, an eastern origin and subsequent westwards spread is a more likely history for M8 than the pincer model suggested by Oppenheimer (2003).

Let us have a look at one last example. The frequency of haplogroup A in East Asia is generally between 5–10% (Yao et al. 2002). Similarly, in Central Asia, it accounts for <10% of the mtDNAs of East Asian origin (Comas et al. 2004). Significantly, only one subclade of A, A4, is present in Central Asia, while A3, A5, A7 and a set of unclassified A* lineages are found alongside A4 in East Asia (see clarification of haplogroup A subgroups classification at www.evolutsioon.ut.ee; cf. Tanaka et al. 2004).

Unfortunately, only a fraction of A4 can be assigned to subclades using the HVS-I motifs of the three completely sequenced examples of A4 (excl. the Native American A2) mtDNAs. Available HVS-I data on Asian populations (e.g. summarized in Metspalu et al. 2004) suggests that haplogroup A4 displays further region-specific subclades. The root HVS-I haplotype has been found mainly in Chinese samples, from both tribal and Han people, but also in tribal populations from East India and Thailand and a few Central Asians and southern Siberians. Transitions at nps 16124, 16260 and 16274 delineate Thai, East Indian and Chinese-specific subclades, respectively, while a number of additional minor branches exist. Most Central Asian and southern Siberian A4 lineages group with the Chinese variants. The observation that the spread of haplogroup A in Central Asia is restricted to only one of the subclades and that,

within that subclade, the lineages present are generally shared with the East Asians is, again, consistent with the eastern origin of haplogroup A.

Together the haplogroups (M8, D, G, and A) shown here to have radiated out from East Asia cover 70 and 80% of the East Eurasian-specific mtDNAs in Central Asia and southern Siberia, respectively. Without going into the details concerning the rest of the maternal lineages of eastern provenance in these regions, we can add that none bear signs of being ancestral to lineages in East Asia.

In their study of Y-chromosome variation, Wells et al. (2001) argued that Central Asia has been the source area of multiple migrations leading both west and east. The metaphor – "important reservoir of genetic diversity" – which was used to describe high genetic diversity among the present-day Central Asians (Wells et al. 2001), is indeed true. However, it appears increasingly more likely (e.g. Comas et al. 2004) that this particular reservoir, as it is reflected in its present gene pool, has been formed by gene flows to Central Asia from both east and west, long after the initial settlement of Asia.

2.2.7. Conclusion

The Southern Coastal Route of the pioneer phase of the peopling of the vast territories of Asia has gained increasingly stronger experimental support, thanks to recently acquired deeper phylogenetic and phylogeographic knowledge about the spread of mtDNA (and Y-chromosomal) variation in this continent. Much, if not all, of the early settlement process can be seen as a "fast train to Southeast Asia and Australia along the SCR" – indeed, so fast that the founder haplotypes at the base of haplogroups M, N, and R (from the two immediate sister groups M and N to a plethora of the sub-Saharan L3 variants) reached all major destinations alongside the route, as far down as Australia. It appears that Central Asia and southern Siberia were not involved in the initial peopling of the Continent. It is also evident that the initial fast train phase has been followed by a long-lasting freezing of the major geographic pools of maternal lineages in the south and further gene flows northwards from Southeast Asia and subsequently back westwards along the Steppe Belt extending from Manchuria to Europe. At present, west Siberia, the Urals and Central Asia form a huge continuous admixture zone encompassing East and West Eurasian maternal lineages—a process that has so far had only a minimal influence to the essentially distinct autochthonous patterns of mtDNA variation in most of South Asia, East Asia, Southeast Asia and Australasia.

2.3. India: Some general considerations

India is a vast and highly heterogeneous region (Figure 4). The current population size exceeds 1 billion and is growing faster than that in China. The major divisions are linguistically based: the most numerous are Indo-European speakers, followed by Dravidian speakers in the south and a smaller number of speakers of Austro-Asiatic (Mundari branch) and Sino-Tibetan languages (see further

Figure 4. Map of India. Shown are major geographic regions and states of India. "*Linguistical data*") dispersed mainly in the eastern parts of the Indian sub-continent.

Many studies on "racial classifications", especially from the first half of 20th century, have been put forward with regard to the origin of the present-day ethnic groups in India. Though, largely contradictory, all agree on the existence of several ethnic groups with distinct morphological features.

The population of India is characterised by the presence of multiple religions. The majority of Indians are Hinduists (82%) and the largest minority religion is Muslim (12%). Other minority religions include Christianity, Buddhism and Jainism, along with the Sikh and Parsi religions.

Hindu society is further dissected by the caste system which classifies its members by birth (Figure 5). The hierarchical system is built up of four main caste groups or *varnas*. The lowest *varna* is called *Sudra* and stands for labourers. People belonging to the other three are collectively called the "twice born". The hierarchically next *varna*, Vysya, includes the merchants, farmers, artisans etc. *Kshatriya* is a common nominator for warriors and rulers whereas at the top of the caste system stands the priests – *Brahmin*. Officially outside the caste system but in fact occupying the lowest rank in the Hindu social hierarchy lie the untouchables – *Pancham.*

Each *varna* is further divided into numerous castes where within caste territorial groups add an extra dimension. Thus, at the end of the day, the Indian caste system is made up of 37 thousand endogamous groups; a phenomenon that has profound implications for population biology. The caste borders are not invincible though. The Hindu system allows what is called upward mobility of women, which means that women of a lower caste can marry into a higher caste.

The most common explanation for the origins of the caste system puts its roots in the Aryan invasion some three to four thousand years ago. This view has, however, been recently criticised by genetic studies which refute the very existence of a substantial invasion of people from the northwest and recognise that the social strata may predate the respective time frame (see e.g. REF III). Also the archaeological grounds for the "single-origin model" of the caste system are turning out to be largely illusionary. In a recent review Nicole Bovin argues that most archaeological studies on the origins of the castes done so far have dealt with the subject in the context of the Indo-Aryan paradigm. She therefore accuses the approach of being biased, largely neglecting the rich archaeological record that leaves ample room for different – multi-origin – models (Bovin 2005). For example the archaeological record seems to contradict the very idea of Indo-Aryan hordes moving into India by masses and imposing a new culture and religion. Rather, the influx appears much more small-scale and the newcomers are seen largely incorporating into existing society (Shaffer 1984; Erdosy 1995; Kennedy 1995; Shaffer and Lichtenstein 1995).

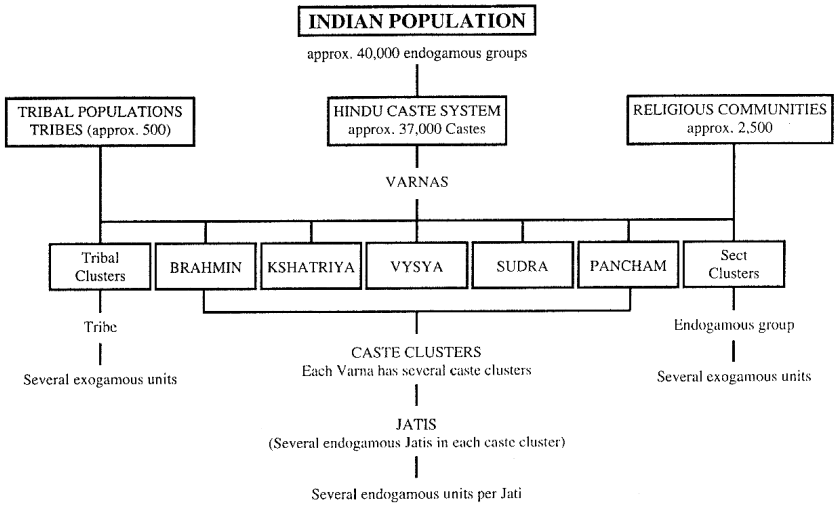

Figure 5. The social organization of the Indian populations. This scheme is further complicated by territorial affiliation of various tribes and caste groups. (Papiha 1996).

In addition to the myriad of castes and religious groups the landscape of the Indian populations is further splintered by the tribal populations. These usually small and endogamous groups are spread over many regions all over India (though not uniformly) accounting for some 8% of the total population size (Figure 6). The tribal populations stand out by exhibiting features not seen among the rest of the populace. Firstly, two out of the four language families spread in India, namely the Austro-Asiatic and Tibeto-Burman, are confined within the tribal groups. Further, hunter-gathering, as the main mode of subsistence, is still practiced by quite a few tribal populations (Singh 1997). Discussion on the origins of the different tribal groups is ongoing. Perhaps the most straightforward is the situation with the Tibeto-Burman speaking tribal populations in Northeast India. In addition to the clear linguistic association with the populations living east of India, the origin from the East is also manifested in the geographic spread of these tribal groups (Figure 9). The situation is much more complex regarding the Austro-Asiatic speaking tribes. Many authors have regarded them as the remnants of the original people of India, while, to the extreme contrast, others postulate more recent roots again east of India. The latter paradigm finds support in the linguistic data as a majority of Austro-Asiatic languages are spread in Southeast Asia while in Peninsular India we see only one branch – the Mundari languages. But note that the Khasi in Meghalaya and Nicobarese in Nicobar Islands speak languages of the other – Mon-Khmer branch.

Taken together the outlined features of the Indian populace render the population structure highly complex. Let us just add that the People of India

project of the Anthropological Survey of India (published in 72 volumes, edited by K.S. Singh see https://www.vedamsbooks.com/asicat.htm for full list of titles) recognises altogether 4635 different ethnic communities.

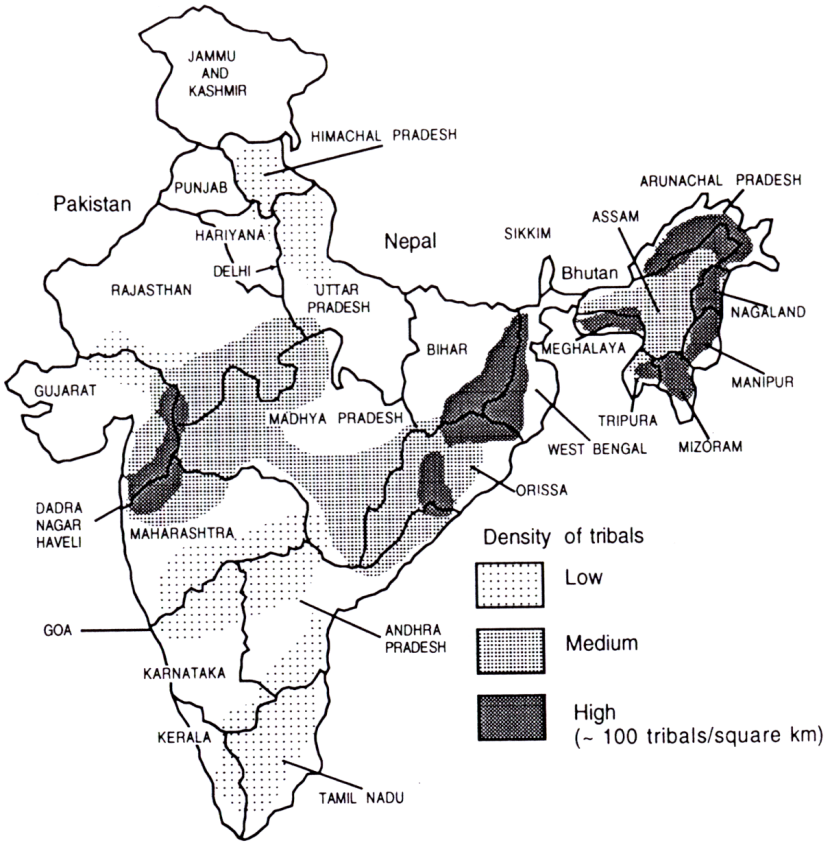

Figure 6. The states of India (before the administrative reform in 2000) with approximate local densities of tribal populations (Cavalli-Sforza et al. 1994).

2.3.2. Palaeontological and archaeological data

As it is the case for the rest of Eurasia, earlier hominid species inhabited India before the immigration of anatomically modern humans. Tool-using *Homo erectus* populations have been in India for over 0.5 MYA. The earliest skeletal evidence comes from a calvarium (the top part of the skull), referred to as the Narmada Man (Sonakia 1984). This fossil is currently attributed to *Homo heidelbergensis* (Cameron et al. 2004) and dated by faunal correlation to ca. 300–250 KYA (Kennedy 2000). Although the Narmada partial cranium is

associated with Late Acheulean (Lower Palaeolithic) artefacts, no hominid remains have been found with Middle Palaeolithic industries. So far the earliest fragmental skeletal evidence (at ca 28 and 33 KYA) of anatomically modern humans come from two caves in Sri Lanka (Kennedy et al. 1987; Kennedy and Deraniyagala 1989). Note that this island was at that time connected with the continent. These fossils are found in association with microlithic industries thus exhibiting the earliest conclusive evidence of both anatomically and behaviourally modern humans in South Asia (James and Petraglia 2005). As discussed before (see Chapter 2.2) it is most likely that modern humans settled in South Asia at least ten thousand years earlier than the time frame of the findings in Sri Lanka. Therefore, although the record of more recent fossils of modern humans in the subcontinent covers the continent reasonably well, in terms of questions relating to the initial settlement of the continent the fossil record of South Asia remains poor.

Understandably the archaeological record in South Asia is much richer. Here we shall not go into detailed description of the lithic industries, which would be a clear overestimation of my expertise. Two recent reviews should be pointed out which give a comprehensive coverage of the subject (Misra 2001; James and Petraglia 2005). Here we briefly touch upon the chronology and interrelations of different stages of the evolution of the Palaeolithic period in South Asia and by the end of the chapter discuss the advent of agriculture in the subcontinent.

The first effective colonization of South Asia was succeeded by hominids practicing the Lower Palaeolithic culture. The exact duration of this period is not fully understood. Misra (2001) concludes that the most likely start point for the Acheulian culture is in the early Pleistocene between 2 and 0.7 MYA and that it persisted well into the late Pleistocene. This means that by the end the Acheulian culture existed side by side with the more advanced Middle Palaeolithic technology. This technology is essentially flake based, consisting of prepared cores, retouched flakes and diminutive bifaces (James and Petraglia 2005). The earliest datings for the Middle Palaeolithic go back to ca. 150 KYA. Whereas the upper limit matches that of the late Pleistocene at 10 KYA (Misra 2001). Blade-based and microlithic industries are characteristic of the next stage of the Palaeolithic technology evolution in South Asia. Many sites of this epoch show an increasing importance of burins and backed tools (tools with one edge blunted). Traditionally this period is called the Upper Palaeolithic after its counterpart in western Eurasia. In their recent review James and Petraglia (in press) argue that these two technologies are essentially different and therefore coin a new term “Late Palaeolithic” to describe what was previously known as “Upper Palaeolithic in South Asia”. Radiocarbon datings put the start of this technology in South Asia generally at 30 KYA (Misra 2001) while the earliest dating for the Late Palaeolithic in the region comes from Pakistan and yields 45 KYA (Dennell et al. 1992). The end of the Late Palaeolithic is concurrent with the end of the Middle Palaeolithic. Therefore, for the whole of its existence the

Late Palaeolithic was coexisting with the preceding technology. Moreover, the considerable temporal and spatial variability of its emergence, as well as the evidence of transitional technologies between the Middle and Late Palaeolithic suggest, that at least some Late Palaeolithic industries developed *in situ* from the preceding Middle Palaeolithic and were not imported from somewhere else (James and Petraglia 2005).

The succeeding Mesolithic technology is primarily based on microliths. A large body of datings for this era places it confidently between 10 and 2 KYA (Misra 2001). Here we note again that the earliest (ca. 30 KYA) microliths in South Asia come from Sri Lanka.

Until recently the origins of agriculture in India were considered explicitly derivative and secondary. A solid body of evidence indicates that the cultivation of plants reached India from two different directions: the Fertile Crescent and East Asia (Diamond and Bellwood 2003). The cradle of agriculture the Near East yielded the agricultural package combining the cultivation of wheat and barley and the domestication of cattle, sheep and goat. Domesticated pigs and rice cultivation in India has been attributed to import from the Yangtze basin. While the "Fertile Crescent" agricultural package reached northwest South Asia nine to seven thousand years ago, it did not spread fast throughout the subcontinent. It reached the Ganges basin and the Deccan in central India little over 4 KYA (Fuller 2003).

Yet, recently this simplistic picture has been challenged. The introduction of the agricultural package of "Fertile Crescent" to Northwest India and its slow but steady spread in the subcontinent is generally not questioned. Nor is the idea of the spread of some elements of early Chinese agriculture into eastern India overthrown. However, in addition to this, local origins for different crop packages are put forward both in South and Northeast India (Figure 8) (Fuller 2003). Archeobotanical evidence suggests that several local varieties of beans (mungbean – *Vigna radiata* and horsegram – *Macrotylora uniflorum*) and millet (browntop millet – *Brachiaria ramosa* and bristley foxtail – *Setaria verticillata*) were domesticated in South India before the "Fertile Crescent" package had any influence on the region (Fuller 2003; Fuller et al. 2004).

Also, in the Ganges valley rice and millets were grown before the South-west Asian crop system arrived. The origins of Indian rice are not fully resolved. The somewhat mainstream idea has been that it was introduced from the east, perhaps together with the Mundari language family (Diamond and Bellwood 2003; Higham 2003). Though it is clear that one occasion of rice domestication took place in East Asia, probably together with the pig and chicken, genetic studies have indicated at least two domestications. It is well possible, then, that the other one took place in the Ganges valley (Fuller 2003). Although Gujarat is situated right beside the Indus valley, which acted as an outpost for the Southwest Asian agricultural package, it seems possible that local domestication of little millets happened here too (Fuller 2003).

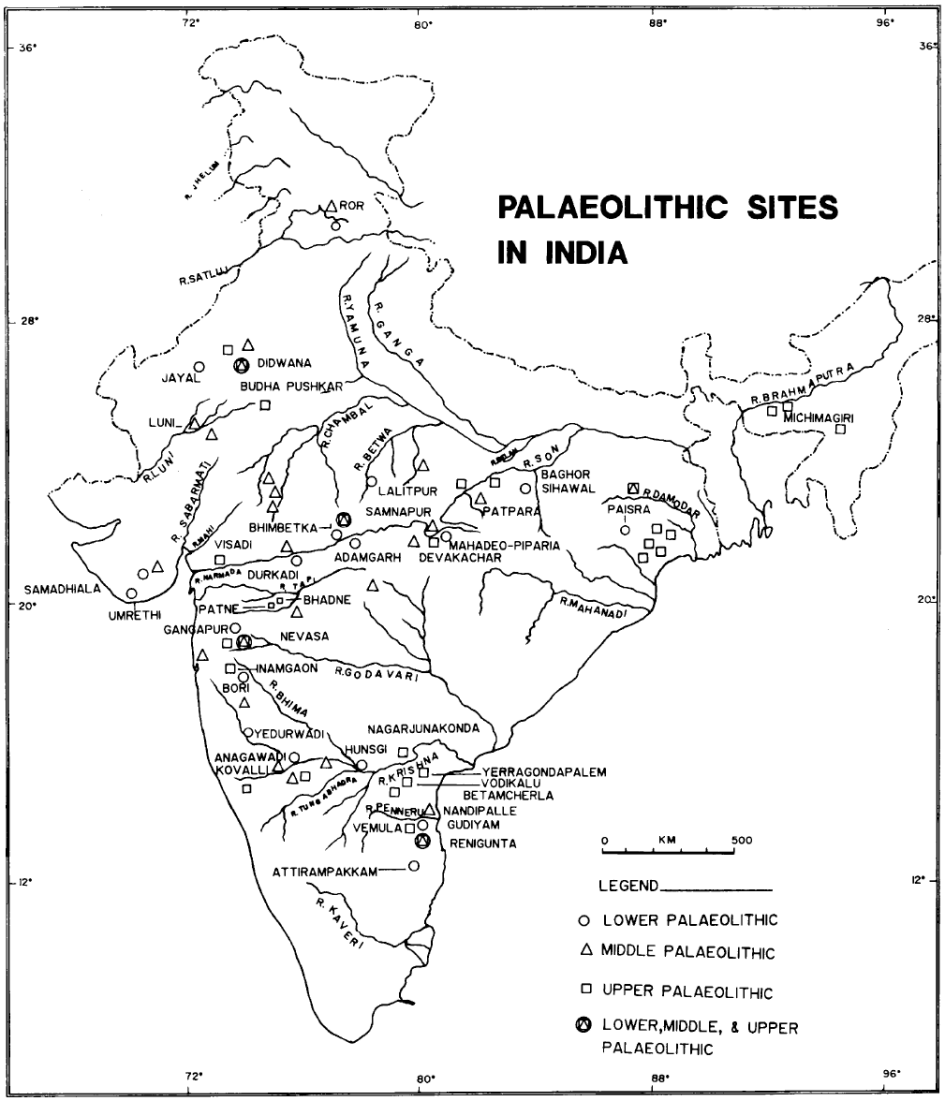

Figure 7. Major Palaeolithic archaeological sites in India (Misra 2001).

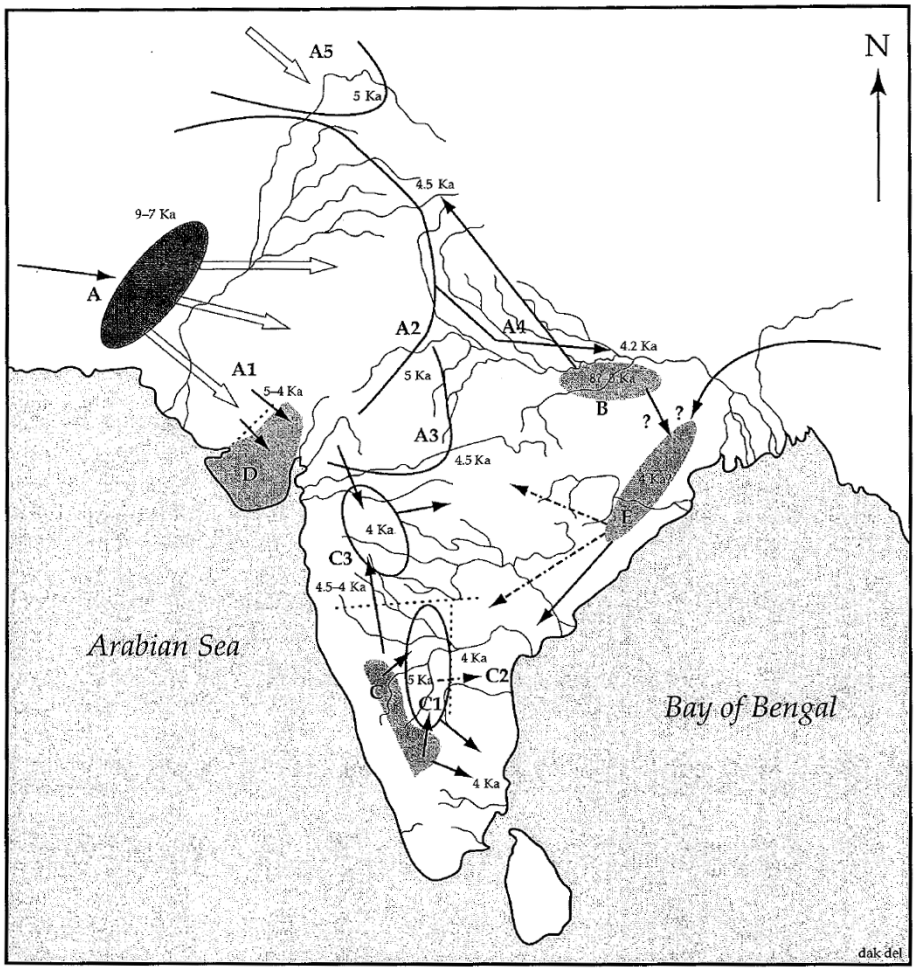

Figure 8. A synthetic view of early agricultural origins and dispersals in South Asia (Fuller 2003). Regions of probable local domestications shaded. Dispersals indicated by arrows and prehistoric agricultural frontiers indicated by lines (solid =”moving”, dotted =”static”). See (Fuller 2003) for full legend.

2.3.3. Linguistic data

Nearly all languages spoken in South Asia today can be assigned to one of four major language families – Austro-Asiatic (Mundari branch), Dravidian, Indo-European and Sino-Tibetan. There are a few, which cannot be assigned to any family, though. Nahali, a tribal language of Central India and Burushaski, spoken by a group of people, the Hunzas, numbering around 40 000 in Pakistan and Afghanistan are two such.

Table 1. Worldwide distribution of the four language families present in India.

Austro-Asiatic	Southeast Asia (Mon Khmer), East and Central India (Mundari)
Dravidian	South and Central India, Pakistan
Indo-European	Europe, West Asia, North, West and East India
Sino-Tibetan	China, Southeast Asia, North and Northeast India

The geographic range of the Austro-Asiatic, Indo-European and Sino-Tibetan speakers is extensive. India harbours only a minority of the languages within these families (Table 1). Dravidian languages are however largely restricted to India. There is only one outlying population – Brahui in Baluchistan (Pakistan). This observation itself has led scientists to come to very different ideas on the origin of the Dravidic languages. Some argue that they might have developed within India (Gadgil 1997) and that the Brahui have migrated north-westward to their present habitat (Elfenbien 1987). Others put the origins of the Dravidic family in West Asia. This line of thought has, in addition to the arguments raised by the present area of the Brahui language, been fuelled by the hypothetical Dravidian affiliation of the extinct Elamic language in the Zagros Mountains of contemporary Iran (McAlpin 1981).

The language-farming dispersal model ties the arrival of southwest Asian agricultural package with that of the Dravidic speakers e.g. (Renfrew 1996). Fuller (2003) makes a strong case for South Asian origins for the Dravidic family. He shows that the Dravidian words for native cultivars such as mung-bean and horsegram are older than those designating the introduced crops of the "Fertile Crescent" agricultural package. On the basis of tree names he further situates proto-Dravidian in the Peninsular tropical dry deciduous/ savannah zone (Dorian Fuller, personal communication).

Most of the Austro-Asiatic speakers (>98%) live in Southeast Asia. In India they are spread sporadically in central and eastern India (but to the west of Bangladesh) and speak languages of the Mundari branch of the Austro-Asiatic family, which is found only in India and further delineates into northern and southern groups (Figure 9). The Khasi in Meghalaya and Nicobarese are the only exceptions representing the Mon Khmer branch. The spread of the Austro-Asiatic languages is quite unequivocally attributed to language-farming dispersal from the Yangtze basin (Diamond and Bellwood 2003; Higham 2003). Analysing the agriculture related vocabulary of the Mundari languages, Fuller (2003) concludes that proto-Mundari reached eastern India after the arrival of the Southwest Asian agricultural package. However, today most Austro-Asiatic speaking communities in India live as hunters-gatherers and/or practice low input shifting cultivation.

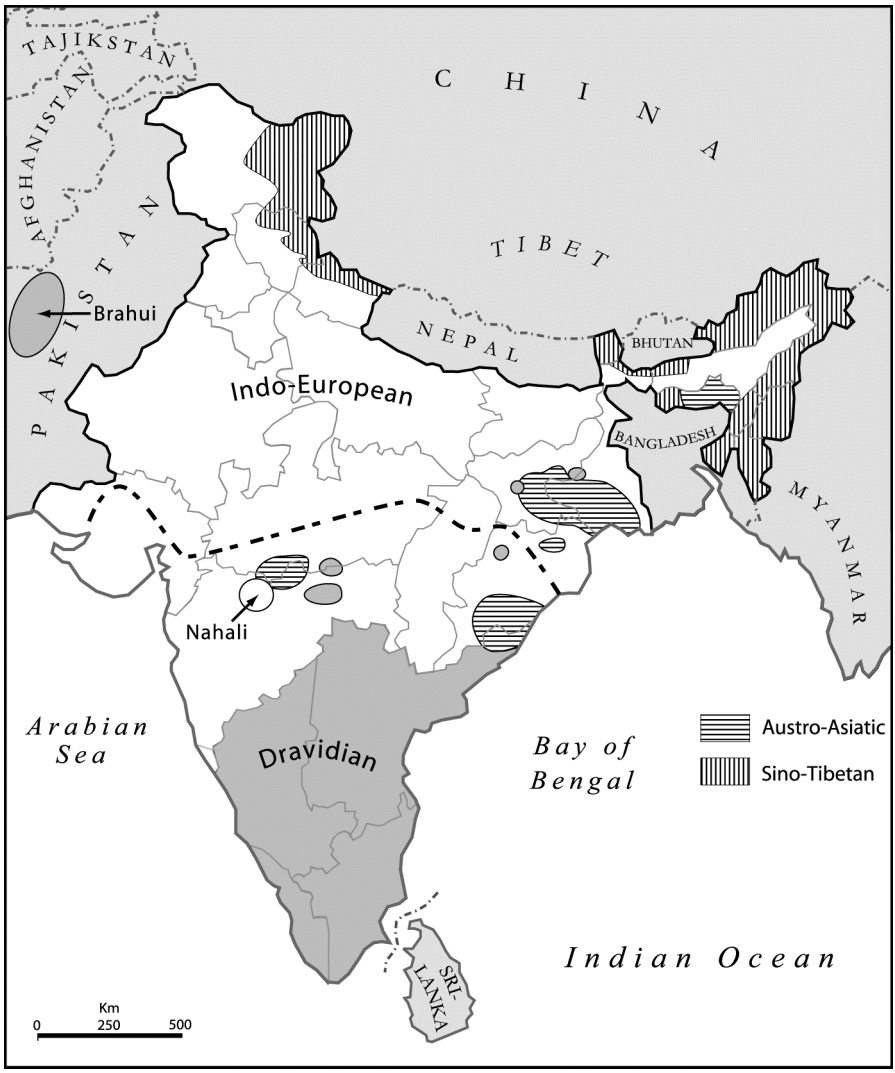

Figure 9. Distribution of the four language families in India. Dashed line indicates general frontier region of former Dravidian linguistic and cultural influence. Adopted from (Fuller 2003) and (Mandal et al. 2002).

Sino-Tibetan speakers of India are also mostly tribal groups, though they also include communities like Maites of Manipur valley practicing advanced agriculture. They live largely along the Himalayas (Figure 9) and many of them report having moved into India from Myanmar or China within the last few generations.

It has to be added here that based on some agricultural vocabulary in extant South Asian languages, which apparently cannot be attributed neither to Mundari, Dravidic, nor Indo-European origin, a possibility for a by now vanished (agricultural) language in the Ganges valley has been proposed (Fuller 2003 and references therein).

Most of the Indian mainland populations are Dravidic and Indo-European speakers. Both include communities of all economic levels from tribals to the most advanced cultivator, pastoral, trader or priestly groups. Many of the technologically less advanced amongst these communities such as Dravidic-speaking Kanis of Kerala or Indo-European-speaking Bhils of Rajasthan, may have acquired these languages in more recent times through the influence of the economically more advanced mainstream societies. It is however notable that while there are several Dravidian speaking forest dwelling tribal communities such as Gonds or Oraons in a matrix of technologically more advanced Indo-European speaking communities, there are no enclaves of forest dwelling tribal Indo-European speakers surrounded by more advanced Dravidic- speaking communities. The tribal Indo-European speakers of South India are all nomadic communities such as Banjaras or Pardhis (Indian Gypsies) with known history of migration from Rajasthan to South India in recent centuries. Some researchers argue that this is strongly suggestive of the Dravidians being older inhabitants of the Indian subcontinent, and that they have been pushed southwards, surrounded by or converted to Indo-European languages by later arriving Indo-European speakers (Gadgil 1997 and references therein). A more northerly than today spread of the Dravidic languages is also supported by the presence of "Dravidic" place names and kinship terminology (Figure 9) (Fuller 2003 and references therein).

2.3.4. Data obtained from studies using "classical" markers

The essence of "classical" ("pre-DNA") genetics lies in the geographical mapping of allele frequencies. The huge amount of data, gathered during the "classical era", allowed to formulate a number of basic problems concerning Indian population genetics – leaving them, *alas*, largely unsolved.

28 Indian populations are included in the monumental study of Cavalli-Sforza and colleagues (Cavalli-Sforza et al. 1994). These cover Dravidian and Indo-European speakers as well as few reasonably well analysed smaller groups (incl. some tribal populations). On a most general level they conclude that in genetic distance trees Indian populations cluster more closely with western Eurasian populations than with other Asians or Africans. In addition they propose that there are at least four major components of the genetic structure of India (Cavalli-Sforza et al. 1994).

This classification is far from being the only one. Another significant summary of the numerous genetic studies on the populations of India is

provided by Surinder S. Papiha (Papiha 1996). He concludes that tribal populations are in general well differentiated from the nontribal castes or communities. Genetic differentiation among nontribal communities and occupational castes is slight, but the subpopulations of each nontribal group of different provinces demonstrate considerable genetic diversity.

It seems fair to say at present that the wealth of knowledge, incorporated into gathered over decades information about the spread of individual classical genetic markers over the subcontinent of India, can be of a significant help in the planning of new research, in particular where DNA-era autosomal markers are involved.

3. AIMS OF THE PRESENT STUDY

When I joined the ICMB evolutionary biology group and started my studies of the maternal legacy of Indian populations, the studies on worldwide human mtDNA variation including that in India were under way (Mountain et al. 1995; Bamshad et al. 1996; Barnabas et al. 1996; Passarino et al. 1996a; Passarino et al. 1996b; Bamshad et al. 1998). The skeleton of global mtDNA phylogeny was emerging (Richards et al. 1998; Torroni et al. 1998; Macaulay et al. 1999b; Quintana-Murci et al. 1999). Yet, basic problems like genetic evidence for the route(s) taken by AMH to colonize Eurasia were not settled. Similarly a number of questions regarding more specific issues of mtDNA diversity and its implications on different (cultural) anthropological hypotheses regarding the prehistory of South Asia needed further attention.

The questions raised for the current work are as follows:

- What is the place of the Indian mtDNA variation on the global mtDNA phylogeny?
- What does the mtDNA variation in India tell us about the peopling of Eurasia?
- Do the Indian tribal populations and the Austro-Asiatic speakers in particular share common ancestry with the castes or are they "more ancient" inhabitants of the subcontinent as suggested by some?
- How extensive was admixture between South Asia and neighbouring regions after the initial settlement? And what can be said about the timescale for the admixture in light of i) suggestions for external origins for all language families present in India, and ii) the advent of agriculture in India?
- In order to tackle the mentioned questions there was an obvious need for an operational enquiry: what is the inner structure of Indian branches of haplogroups M and N(mostly present as lineages arising from the R node)

New papers on mtDNA phylogeography are published nearly weekly. Therefore, after the publication of the papers that are the basis of this thesis a number of reports on mtDNA diversity in South Asia have appeared (Baig et al. 2004; Palanichamy et al. 2004; Wooding et al. 2004; Banerjee et al. 2005; Barnabas et al. 2005; McElreavey and Quintana-Murci 2005; Rajkumar et al. 2005; Sharma et al. 2005; Thangaraj et al. 2005a; Thangaraj et al. 2005b). These papers are not discussed in the current thesis but it has to be added that the results presented in them, though enriching the overall knowledge base about the diversity of the Indian maternal heritage, do not refute any of the conclusions of the present study.

4. RESULTS AND DISCUSSION

For the current thesis mtDNA sequence variation was studied in a total of 796 Indian samples, most of which are held in a collection at Newcastle University (Papiha 1996), and 463 Iranian samples of the EBC collection. For Ref IV the new Indian mtDNA sequence data was combined with that previously published on Indian populations to produce a pooled dataset of 2572 samples. In many cases the published data was reanalyzed. A reduced median network approach was used to reconstruct the topology of mtDNA haplogroups (Bandelt et al. 1995; Bandelt et al. 1999). The coalescence estimates were obtained using estimator ρ, taking 20,180 years as an average time for the fixation of one transition between nps 16090–16365 (Forster et al. 1996). Detailed descriptions of methods and studied populations are given in references I–IV.

4.1. A deep link between South and West Eurasian maternal lineages (Ref I)

All the mtDNA lineages found in Indians fall within the two Eurasian founder maco-haploroups, M and N, which stem from the African super haplogroup L3. This observation supports the recent African origins for modern humans and the replacement of any pre-existing hominid species in the Indian sub-continent.

MtDNA variation among the Dravidic and Indo-European speaking Indians is built up from the same stock of mtDNA lineages. There are no mtDNA haplogroups specific to either language group. Only a small fraction of the Indian mtDNA pool has relatively recent roots in western Eurasia, undermining the hypotheses for large-scale invasion of the Aryans some four KYA.

Haplogroup U provides an interesting link between Indians and West Eurasians. It is the second most frequent mtDNA linage cluster in both regions. Importantly, most of U lineages in India coalesce to a founder haplotype U2 which dates back to over 50 KYA. This estimate is similar to that for the European-specific sub-haplogroup U5.

4.2. The caste and tribal populations of India share a common ancestry in Late Pleistocene (Ref II)

The analysis of mtDNA variation among two tribal populations – Chenchus and Koyas from southern India, placed into context of other tribal and caste populations of India, suggests that the extant Indian caste and tribal populations derive largely from a common mtDNA pool that diversified within India in the

Late Pleistocene. It is also most likely that this ancient population, the carrier of this pool, constituted the initial settlers of the subcontinent.

20 of the 18 and 32 individual mtDNA HVSI haplotypes found amongst the 96 Chenchu and 81 Koya samples, respectively, were also found among other caste and tribal populations of India (total n = 1093). Whereas all but one of the haplotypes that were unique to either of the two tribal populations had one- or two-mutational-step neighbours in the total Indian sample. In contrast, a search carried out by us of the western Asian database consisting of 1,232 samples revealed only two unspecific HVS-I matches with Chenchu and Koya HVS-I haplotypes.

All the mtDNA variants of the Chenchus and Koyas fell within the two Eurasian founder macro-haplogroups M and N (including R) (Ref II. Table 1, Fig 1). Genetic drift is more powerful in small endogamous populations resulting in lower diversity. This was manifested in both of the tribal groups in this study – especially so in the Chenchus where all but three samples belonged to hg M. Also, haplotype diversity in Chenchus was low (0.87; Ref II. Table 1). A third of the Koya mtDNAs belonged to unclassified lineages of haplogroup R, while only one fell into haplogroup U. The low frequency of haplogroup U and the absence of western Eurasian-specific haplogroups HV, JT, N1 and X differentiates these tribes from the Indian caste populations. The picture is rather similar when one looks to tribal populations in other Indian states, e.g., West Bengal and Tamil Nadu.

The most frequent M subhaplogroup among Chenchus and Koyas is the newly defined hg M2 (mtDNA coding region polymorphisms (447G, 1780, and 8502) (Ref II. Fig. 2, Table 2). This subhaplogroup is the most ancient (73,000±22,900 YBP) amongst mtDNA haplogroups, identified so far in Indian populations. M2 is also frequent among the caste populations in southern India, but appears to be completely absent outside India – a highly informative observation, strongly suggesting the lack of any significant migrations of people out of India after the beginning of the initial *in situ* diversification of the first colonizers.

South Asia exhibits a vast variety of autochthonous lineages of all three Eurasian founder haplogroups – M, N and R, whereas regions west of India lack hg M branches (other than M1). This is consistent with the scenario that the SCR migration(s) from Africa carried those variants ancestral to Eurasian mtDNA first along the southern coast towards the Indian subcontinent. A somewhat reduced package of mtDNA variants has reached western Asia and Europe probably somewhat later, though still at the end of Middle Palaeolithic – beginning of Upper Palaeolithic (see Fig 2 of this theses) This model is also supported by the distribution pattern of major Y-chromosome lineages (the reduction of the package of C, D, F, and K in South Asia to F and K in western Eurasia).

After the initial colonisation, region specific mtDNA variants of the founders started to emerge all over the continent. Some of these branches, arose

outside South Asia, (e.g., the western Asian HV and TJ lineages) have, via continuous or episodic low-level gene flow, reached back to India.

4.3. Post Pleistocene spread of languages and farming into South Asia shaped the pre-existing mtDNA pool only marginally (Ref III)

The vast number of different languages spoken today in South Asia belong to four language families. Most numerous are speakers of Indo-European languages in northern India. Dravidic languages are spoken in the south (except the Brahui in Pakistan) while languages belonging to Tibeto-Burman and Austro-Asiatic (predominantly Mundari branch) families are spoken exclusively by tribal populations in the east and northeast of the subcontinent. Supposedly all four language families have been introduced into India rather recently (e.g. Renfrew 1996; Diamond and Bellwood 2003). The spread of the Indo European languages has most often been connected with the widely spread hypothesis of Indo-Arian invasion some 4000 YBP. Arising from the hypothesised common origin between the Dravidian languages and the long vanished Elamic language in Iran, the spread of the former in India has been attributed to migration from West Asia, possibly together with the spread of agriculture from the Fertile Crescent (the "agricultural package" of wheat, barley, cattle, sheep and goat domestication). On the other hand, a South Indian origin for the Dravidic languages has been also suggested with an increasing vigour. Another agricultural package combining cultivation of rice and the domestication of pig, has reached South Asia likely from the northeast – from its origin in the Yangtze valley and the spread of the Austro-Asiatic family is often tied to this dispersal. Recent origins from the northeast are also claimed for the Tibeto-Burman languages. Thus, if these hypotheses hold to the extreme, all the pre Holocene languages in India have been vanished by replacement. To what extent, does this "permanent and total (multiple?) replacement" scenario reflect the phylogeographic and temporal characteristics of the mtDNA pool of the extant South Asians? Not too much indeed. The vast majority of the present-day Indians carry mtDNAs belonging to autochthonous Indian haplogroups that coalesce far deeper than the introduction of the mentioned languages and either of the agriculture packages. Only about eight per cent of the Indian mtDNA pool can be ascribed to more recent – Neolithic and later – admixture from the West (Ref III. Table 17.4). What can be said about this 'putative recent import from the west' (PRIFW) component of the Indian mtDNA pool? Compared to the composition of the mtDNA pool of Europeans, the PRIFW component in Indians differs in higher frequencies of HV and U1 and lower frequencies of H and U5 – a pattern similar to that observed for Anatolians and Iranians. In turn, comparing to West Asians the Indian PRIFW lineages nearly lack U3 and K but

have higher frequencies of I and U4. In the European mtDNA pool haplogroups T1, U3 and J are often considered as the signals of the Neolithic spread of farmers from the Near East. It is intriguing, that the Indian PRIFW lineages show a relatively lower share of those haplogroups when compared to Iran (compare a T1 to T ratio of 2/ 14 in India versus15/38 in Iran). It is possible, then, that Iranians obtained most of their U3, K, J, Tl and also X lineages only after a substantial diffusion of Proto-Iranian' lineages to the Indian mtDNA pool had taken place. Taken together it seems that a notable part of the eight per cent of the PRIFW mtDNA lineages in India may actually predate the proposed introduction of languages and agriculture listed above.

4.4. The extant mtDNA pool of South Asia was inoculated during the initial peopling of the continent and has not been replaced since (Ref IV)

Most contemporary Indians trace their maternal roots in the M macro-haplogroup. The frequency of hg M is somewhat higher among the tribal populations (72%) than among the caste populations (58%). It is important to note that the frequency of hg M drops abruptly west from India. In Iran, hg M constitutes a mere 5% of the mtDNA gene pool. Moreover, less than half of it can be ascribed to geneflow from India, since the rest affiliates clearly with M subhaplogroups specific to either East Asia or northwest Africa. There are no Iranian- specific haplogroup M lineages.

Contrary to that, the frequency of haplogroup M stays high when one moves to the east of India. However, the important aspect of this continuity is that there is no overlap between haplogroup M subhaplogroups specific to India and East Asia, suggesting unambiguously that the most frequent mtDNA lineages, present in the Indian mtDNA pool, have never spread outside India after the very initial settlement of their carriers in the subcontinent

Three Indian-specific haplogroups, M2, U2(comprising of sub-hgs U2a,b and c) and R5, which encompass about 15% of the Indian mtDNA pool, exhibit equally deep coalescence ages of about 50,000–70,000 years. These haplogroups are the most ancient autochthonous mtDNA haplogroups in India, therefore their spread can be associated with the population expansion upon the initial peopling of South Asia. It has to be noted, however, that this list of “ancient haplogroups” will likely expand. As a result of increasing resolution of the South Asian mtDNA phylogeny, new subhaplogroups may still emerge from the forest of yet unclassified haplogroup M and N(R) lineages.

Haplogroups U7, W and R2 harbour a number of similar traits. Their overlapping geographic distribution and comparable coalescence times suggest some degree of genetic continuum in the area spanning from the Near and Middle East, through northwest India and reaching north into Central Asia. The coales-

cence estimates for these haplogroups are equally deep (around 30,000–50,000 years) in these different regions. Because the South- and West Asian-specific sub-branches of haplogroup U7 predate the LGM , the phylogeography of these haplogroups is better explained by a deep autochthonous history than by recent admixture.

Through the use of mtDNA coding region markers, we were able to classify altogether a quarter of the Indian M and R mtDNAs into a number of Indian-specific mtDNA haplogroups, four of which have been identified firstly by us. Several of such monophyletic clusters are characterized by clear patterns in their geographic distribution, which sometimes translate into different frequencies among different socio-cultural groups of India. For example, haplogroup M6 is primarily spread along the western shore of the Bay of Bengal where its subclades, M6a and M6b, keep apart in the southwest and in the northeast, respectively. In contrast, the spatial frequency map (Ref IV, Fig 1) of haplogroup M3a demonstrates its frequency radiation out of Rajasthan (with a significant cline, see Ref IV, Fig 4). Due to the increased frequency towards the southern part of India (Ref IV, Fig 1), haplogroup M2 is significantly ($p < 0.05$) more frequent among the Dravidic speakers than among the Indo-European speakers, who are spread mostly in the northern regions of India (Ref IV, Table 2). It is more plausible that geography rather than linguistics is behind this pattern, because the frequency of M2 amongst the Indo-European speaking populations in southern India is significantly higher than that in the north, while there is no significant difference between Dravidic and Indo-European speaking populations from the same geographic region (Ref IV, Table 2).

We conclude that the initial mtDNA pool established upon the peopling of South Asia has not been replaced but has rather been reshaped in situ by major demographic episodes in the past and garnished by relatively minor events of gene flow both from the West and the East during more recent chapters of the demographic history in the region.

5. CONCLUSIONS

1. The extant Indian caste and tribal populations, covering speakers of different language groups, derive largely from a common source population that diversified within India in the Late Pleistocene. Only the Tibeto-Burman speakers in eastern and northern India show their more recent maternal ancestry in East Asia.
2. All the mtDNA variants of Indian populations, including the tribal populations like the Chenchus and Koyas, belong to the two Eurasian founder macro-haplogroups M and N (including R) although frequencies of individual haplogroups may vary substantially between (drift-prone tribal) populations.
3. The most frequent and at the same time most ancient (73 ±22.9 KYA as the signal of the beginning of its expansion) hg M subclade among Indian populations is the newly defined haplogroup M2 (mtDNA coding region polymorphisms 447G, 1780, and 8502).
4. Three Indian-specific haplogroups, M2, U2a,b and c and R5, which encompass about 15% of the Indian mtDNA pool, exhibit equally deep coalescence ages at about 50–70 KYA. Thus, their spread can be associated with the initial peopling of South Asia.
5. We made progress in reconstructing the substructures of haplogroups M and R identifying new Indian-specific clusters, several of which show clear pattern in their spread within the subcontinent.
6. Only a minor part of the extant Indian mtDNA pool can be ascribed to the admixture from the West during the Holocene. This is not in agreement with putative large scale immigrations from the West i.e. the Indo-Aryan invasion. Because genes and cultural traits may or may not migrate together, the possibility for exterior origins for all the existing language families in India cannot be ruled out, but local origin for the Dravidic family appears more parsimonious.
7. We found that haplogroup M frequency drops abruptly from about 60% in India to about 5% in Iran, marking the western border of the haplogroup M distribution. A similarly sharp border cuts the distribution of Indian-specific mtDNA haplogroups to the east and to the north of the subcontinent. Our results lend further credence to the hypothesis that the initial mtDNA pool established upon the peopling of South Asia has evolved and been constantly reshaped *in situ* by numerous pre-historic and historic demographic episodes, while being garnished by relatively minor events of gene flow both from the West and the East during more recent chapters of the demographic history in the region.
8. The presence of M and N (including R) basal lineages in South Asia and further east, together with the lack of hg M in West Eurasia, is consistent with the single Southern Coastal Route of peopling Eurasia.

REFERENCES

Adams JM, Faure H (1997) Global land environments during the last interglacial. Oak Ridge National Laboratory, TN, USA

Adams JM, Maslin M, Thomas E (1999) Sudden climate transitions during the Quaternary. Prog Phys Geogr 23:1–36

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147

Avise JC (2000) Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, Massachusetts

Awadalla P, Eyre-Walker A, Smith JM (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science 286:2524–2525

Baig MM, Khan AA, Kulkarni KM (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

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, Wallace DC (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics 130:139–152

Bamshad M, Fraley AE, Crawford MH, Cann RL, Busi BR, Naidu JM, Jorde LB (1996) mtDNA variation in caste populations of Andhra Pradesh, India. Hum Biol 68:1–28

Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BV, Rasanayagam A, Hammer MF (1998) Female gene flow stratifies Hindu castes. Nature 395:651–652

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753

Bandelt HJ, Kong QP, Parson W, Salas A (2005) More evidence for non-maternal inheritance of mitochondrial DNA? J Med Genet

Bandelt HJ, Salas A, Lutz-Bonengel S (2004) Artificial recombination in forensic mtDNA population databases. Int J Legal Med

Banerjee J, Trivedi R, Kashyap VK (2005) Mitochondrial DNA control region sequence polymorphism in four indigenous tribes of Chotanagpur plateau, India. Forensic Sci Int 149:271–274

Barker G, Barton H, Beavitt P, Bird M, Daly P, Doherty C, Gilbertson D, Hunt C, Krigbaum J, Lewis H, Manser J, McLaren S, Paz V, Piper P, Pyatt B, Rabett R, Reynolds T, Rose J, Rushworth G, Stephens M (2002) Prehistoric foragers and farmers in Southeast Asia: renewed investigations at Niah Cave, Sarawak. Proc Prehistoric Soc 68:147–164

Barnabas S, Apte RV, Suresh CG (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 (2005) High-Resolution mtDNA Studies of the Indian Population: Implications for Palaeolithic Settlement of the Indian Subcontinent. Annals of Human Genetics epubl

Bar-Yosef O (2001) Dating the transition from the Middle to Upper Palaeolithic. Paper presented at XXI[e] Rencontres internationales d'archeologie et d'historie d'Antibes. Antibes

Bermisheva MA, Kutuev IA, Korshunova TY, Dubova NA, Villems R, Khusnutdinova EK (2004) Phylogeographic analysis of mitochondrial DNA in the Nogays: A strong mixture of maternal lineages from eastern and western Eurasia. Mol Biol (Mosk) 38:516–523

Bovin N (2005) Orientalism, ideology and identity: Examining caste in South Asian archaeology. Journal of Social Archaeology 5:225–252

Bowler J, Johnston H, Olley J, Prescott J, Roberts R, Shawcross W, Spooner N (2003) New ages for human occupation and climatic change at Lake Mungo, Australia. Nature 421:837–840

Bromham L, Eyre-Walker A, Smith NH, Smith JM (2002) Mitochondrial Steve: paternal inheritance of mitochondria in humans. TRENDS in Ecology and Evolution 18:2–4

Brown WM, George M, Jr., Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci U S A 76:1967–1971

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

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, N.J.

Chakrabarti DK (1999) India. An archaeological history. Palaeolithic beginnings to early historic foundations. Oxford University Press, New Delhi

Chappell J (2002) Sea level changes forced ice breakouts in the Last Glacial cycle: new results from coral terraces. Quaternary Sci Rev 21:1229–1240

Chen TM, Zhang YY (1991) Palaeolithic chronology and possible co-existence of *H. erectus* and *H. sapiens* in China. World Archeol 23:147–154

Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, Bertranpetit J (2004) Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. Eur J Hum Genet 12:495–504

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R (2004) Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. Am J Hum Genet 74:1014–1022

D'Aurelio M, Gajewski CD, Lin MT, Mauck WM, Shao LZ, Lenaz G, Moraes CT, Manfredi G (2004) Heterologous mitochondrial DNA recombination in human cells. Hum Mol Genet 13:3171–3179

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

Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Miscicka-Sliwka D, Wozniak M, Zakharov IA (2003) Diversity of mitochondrial DNA lineages in South Siberia. Ann Hum Genet 67:391–411

Diamond J, Bellwood P (2003) Farmers and their languages: the first expansions. Science 300:597–603
Dolukhanov PM, Shukurov AM, Tarasov PE, Zaitseva GI (2002) Colonization of northern Eurasia by modern humans: radiocarbon chronology and environment. J Arch Sci 29:593–606
Egger J, Wilson J (1983) Mitochondrial inheritance in a mitochondrially mediated disease. N Engl J Med 309:142–146
Elfenbien J (1987) The periplus of the "Brahui Problem". Studia Iranica 16:215–233
Elson JL, Samuels DC, Turnbull DM, Chinnery PF (2001) Random intracellular drift explains the clonal expansion of mitochondrial DNA mutations with age. Am J Hum Genet 68:802–806.
Elson JL, Turnbull DM, Howell N (2004) Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. Am J Hum Genet 74:229–238
Erdosy G (1995) Language, Material Culture and Ethnicity: Theoretical Perspectives. In: Erdosy G (ed) The Indo-Aryans of Ancient South Asia,. Munshiram Manoharlal, New Delhi, pp 1–31
Etler DA (1996) The fossil evidence for human evolution in Asia. ANNU REV ANTHROPOL 25:275–301
Eyre-Walker A, Smith NH, Smith JM (1999) How clonal are human mitochondria? Proc R Soc Lond B Biol Sci 266:477–483
Filosto M, Mancuso M, Vives-Bauza C, Vila MR, Shanske S, Hirano M, Andreu AL, DiMauro S (2003) Lack of paternal inheritance of muscle mitochondrial DNA in sporadic mitochondrial myopathies. Ann Neurol 54:524–526
Foley R (1998) The context of human genetic evolution. Genome Res 8:339–347
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, Torroni A, Renfrew C, Röhl A (2001) Phylogenetic star contraction applied to asian and Papuan mtDNA evolution. Mol Biol Evol 18:1864–1881
Fuller D (2003) An Agricultural Perspective on Dravidian Historical Linguistics: Archaeological Crop Packages, Livestock and Dravidian Crop Vocabulary. In: Bellwood P, Renfrew C (eds) 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
Gadgil M, Joshi, N.V., Shambu Prasad,U.V., Manoharan,S., Suresh, Patil (1997) Peopling of India. In: Rao BaNA (ed) The Indian Human Heritage. Universities Press, Hyderabad, India, pp pp.100–129
Gilead I (1991) The Upper Palaeolithic period in the Levant. J World Prehist 5:105–154
Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci U S A 77:6715–6719
Greenberg BD, Newbold JE, Sugino A (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21:33–49
Hagelberg E (2003) Recombination or mutation rate heterogeneity? Implications for Mitochondrial Eve. Trends Genet 19:84–90

Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhovel W, Clegg JB, Bowden DK (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. Proc R Soc Lond B Biol Sci 266:485–492

Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhovel W, Clegg JB, Bowden DK (2000) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. Proc R Soc Lond B Biol Sci 267:1595–1596

Hey J (2000) Human mitochondrial DNA recombination: can it be true? Trends in Ecology and Evolution 15:181–182

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

Ho SY, Phillips MJ, Cooper A, Drummond AJ (2005) Time Dependency of Molecular Rate Estimates and Systematic Overestimation of Recent Divergence Times. Mol Biol Evol 22:1561–1568

Ingman M, Gyllensten U (2001) Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. J Hered 92:454–461.

Ingman M, Gyllensten U (2003) Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. Genome Res 13:1600–1606

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

Innan H, Nordborg M (2002) Recombination or mutational hot spots in human mtDNA? Mol Biol Evol 19:1122–1127

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

Jenuth JP, Peterson AC, Fu K, Shoubridge EA (1996) Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nat Genet 14:146–151.

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

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

Jorde LB, Bamshad M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931.

Joshi RV (1996) SOUTH ASIA in the period of *Homo sapiens neanderthalensis* and contemporaries (Middle Palaeolithic). Vol I. UNESCO

Kennedy K (1995) Have Aryans been Identified in the Prehistoric Skeletal Record from South Asia? Biological Anthropology and Concepts of Ancient Races. In: Erdosy G (ed) The Indo-Aryans of Ancient South Asia,. Munshiram Manoharlal, New Delhi, pp 33–66

Kennedy KA, Deraniyagala SU, Roertgen WJ, Chiment J, Disotell T (1987) Upper Pleistocene fossil hominids from Sri Lanka. Am J Phys Anthropol 72:441–461.

Kennedy KAR (2000) God Apes and Fossil Men: Palaeoanthropology in South Asia. University of Michigan Press, Ann Arbor

Kennedy KAR, Deraniyagala SU (1989) Fossil remains of 28,000-year-old hominids from Sri Lanka. Curr Anthropol 30:394–399

Kivisild T (2000) PhD Thesis: The Origins of Souhern and Western Eurasian Populations: an mtDNA Study. PhD, Tartu University, Tartu

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, 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 SS, 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 SS, Deka R, Chakraborty R (eds) Genomic diversity. Kluwer Academic/Plenum Publishers, pp 135–152

Kivisild T, Papiha SS, Rootsi S, Parik J, Kaldma K, Reidla M, Laos S, Metspalu M, Pielberg G, Adojaan M, Metspalu E, Mastana SS, Wang Y, Gölge M, Demirtas H, Schnekenberg E, Stefano GF, Geberhiwot T, Claustres M, Villems R (2000) An Indian ancestry: a key for understanding human diversity in Europe and beyond. In: Renfrew C, Boyle K (eds) Archaeogenetics: DNA and the population prehistory of Europe. McDonald Institute for Archaeological Research University of Cambridge, Cambridge, pp 267–279

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, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk H-V, Stepanov V, Gölge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet 72:313–332

Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis KK, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knjiff P, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2005) The role of selection in the evolution of human mitochondrial genomes. Genetics

Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha SS, Bandelt H-J, Villems R (2002) The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol 19:1737–1751 (erratum 1720:1162)

Kivisild T, Villems R (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931

Klein R (1999) The Human Career. Univ. Chicago Press, Chicago

Koehler CM, Lindberg GL, Brown DR, Beitz DC, Freeman AE, Mayfield JE, Myers AM (1991) Replacement of bovine mitochondrial DNA by a sequence variant within one generation. Genetics 129:247–255.

Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, Zhang Y-P (2003) Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. Am J Hum Genet 73:671–676

Kraytsberg Y, Schwartz M, Brown TA, Ebralidse K, Kunz WS, Clayton DA, Vissing J, Khrapko K (2004) Recombination of human mitochondrial DNA. Science 304:981

Kumar S, Hedrick P, Dowling T, Stoneking M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931.

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

Kuzmin YV, Keates SG (2004) Comment on "Colonization of northern Eurasia by modern humans: radiocarbon chronology and environment" by P.M. Dolukhanov,

A.M. Shukurov, P.E. Tarasov and G.I. Zaitseva. Journal of Archaeological Science 29, 593–606 (2002). J Arch Sci 31:141–143
Lahr M, Foley R (1994) Multiple dispersals and modern human origins. Evol Anthropol 3:48–60
Leonard WR, Sorensen MV, Galloway VA, Spencer GJ, Mosher MJ, Osipova L, Spitsyn VA (2002) Climatic influences on basal metabolic rates among circumpolar populations. Am J Hum Biol 14:609–620
Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, R.J. H (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Am J Hum Genet 74:532–544
Lutz S, Weisser HJ, 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
Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2:13
Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt HJ, Oppenheimer S, Torroni A, Richards M (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science 308:1034–1036
Macaulay V, Richards M, Sykes B (1999a) Mitochondrial DNA recombination – no need to panic. Proc R Soc Lond B Biol Sci 266:2037–2039
Macaulay VA, Richards MB, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B, Torroni A (1999b) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64:232–249
Mandal H, Mukheje S, Datta A (2002) India An Illustrated Atlas of Tribal World. Athropological Survey of India, Calcutta
Marchington DR, Macaulay V, Hartshorne GM, Barlow D, Poulton J (1998) Evidence from human oocytes for a genetic bottleneck in an mtDNA disease. Am J Hum Genet 63:769–775
Maruyama S, Minaguchi K, Saitou N (2003) Sequence polymorphisms of the mitochondrial DNA control region and phylogenetic analysis of mtDNA lineages in the Japanese population. Int J Legal Med 117:218–225
Mathiesen C, Hagerhall C (2002) Transmembrane topology of the NuoL, M and N subunits of NADH:quinone oxidoreductase and their homologues among membrane-bound hydrogenases and bona fide antiporters. Biochim Biophys Acta 1556:121–132
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
McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 160:1231–1241
Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry ST, Wallace DC (1991) The structure of human mitochondrial DNA variation. J Mol Evol 33:543–555

Merriwether DA, Hodgson JA, Friedlaender FR, Allaby R, Cerchio S, Koki G, Friedlaender JS (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 DM, Gilbert MTP, Endicott P, Mastana S, Papiha SS, 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

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA 100:171–176

Misra VN (2001) Prehistoric human colonization of India. J Biosci 26:491–531

Moilanen JS, Finnila S, Majamaa K (2003) Lineage-specific selection in human mtDNA: lack of polymorphisms in a segment of MTND5 gene in haplogroup J. Mol Biol Evol 20:2132–2142

Moilanen JS, Majamaa K (2003) Phylogenetic network and physicochemical properties of nonsynonymous mutations in the protein-coding genes of human mitochondrial DNA. Mol Biol Evol 20:1195–1210

Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, Cavalli-Sforza LL (1995) Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56:979–992

Olivo PD, Van de Walle MJ, Laipis PJ, Hauswirth WW (1983) Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. Nature 306:400–402

Olszewski DI, Dibble HL (1994) The Zagros Aurignacian. Curr Anthropol 35:68–75

Oppenheimer S (2003) Out of Eden: the peopling of the world. Constable, London

Pakendorf B, Stoneking M (2005) Mitochondrial DNA and human evolution. Annu Rev Genomics Hum Genet 6:165–183

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

Papiha SS (1996) Genetic variation in India. Hum Biol 68:607–628

Passarino G, Semino O, Bernini LF, Santachiara-Benerecetti AS (1996a) Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. Am J Hum Genet 59:927–934

Passarino G, Semino O, Modiano G, Bernini LF, Santachiara Benerecetti AS (1996b) mtDNA provides the first known marker distinguishing proto-Indians from the other Caucasoids; it probably predates the diversification between Indians and Orientals. Ann Hum Biol 23:121–126

Piganeau G, Eyre-Walker A (2004) A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. Heredity 92:282–288.

Poulton J, Macaulay V, Marchington DR (1998) Mitochondrial genetics '98 is the bottleneck cracked? Am J Hum Genet 62:752–757

Quintana-Murci L, Chaix R, Wells S, Behar D, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti A, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Mehdi Q, 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
Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat Genet 23:437–441
Rajkumar R, Banerjee J, Gunturi HB, Trivedi R, Kashyap VK (2005) Phylogeny and antiquity of M macrohaplogroup inferred from complete mt DNA sequence of Indian specific lineages. BMC Evol Biol 5:26
Ray N, Adams JM (2001) A GIS-based vegetation map of the world at the Last Glacial Maximum (25,000–15,000 BP). Internet Archaeology 11
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
Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251–1276
Richards MB, Macaulay VA, Bandelt H-J, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62:241–260
Roberts RG, Jones R, Smith MA (1990) Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. Nature 345:153–156
Roberts RG, Jones R, Spooner NA, Head MJ, Murray AS, Smith MA (1994) The human colonisation of Australia: optical dates of 53,000 and 60,000 years bracket human arrival at Deaf Adder Gorge, Northern Territory. Quaternary Science Reviews 13:575–583
Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. Science 303:223–226.
Sauer C (1962) Seashore – primitive home of man? Proc Am Phil Soc 106:41–47
Schwartz M, Vissing J (2002) Paternal inheritance of mitochondrial DNA. N Engl J Med 347:576–580.
Schwartz M, Vissing J (2003) New patterns of inheritance in mitochondrial disease. Biochem Biophys Res Commun 310:247–251
Schwartz M, Vissing J (2004) No evidence for paternal inheritance of mtDNA in patients with sporadic mtDNA mutations. J Neurol Sci 218:99–101
Scott GR, Turner CGI (1997) The anthropology of modern human teeth. Dental morphology and its variation in recent human populations. Cambridge University Press
Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74:1023–1034
Shaffer J (1984) The Indo-Aryan Invasions: Cultural Myth and Archaeological Reality. In: Lukacs J (ed) The People of South Asia: The Biological Anthropology of India, Pakistan and Nepal. Plenum, London, pp 77–90
Shaffer J, Lichtenstein D (1995) The Concepts of “Cultural Tradition” and “Palaeoethnicity” in South Asian Archaeology’. In: Erdosy G (ed) The Indo-
Aryans of Ancient South Asia,. Munshiram Manoharlal, New Delhi, pp 126–154

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

Shen G, Wang W, Wang Q, Zhao J, Collerson K, Zhou C, Tobias PV (2002) U-Series dating of Liujiang hominid site in Guangxi, southern China. J Hum Evol 43:817–829

Singh KS (ed) (1997) The Scheduled Tribes. Vol III. Oxford University Press, Oxford

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

Stoneking M, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics 124:717–733

Stringer C (2000) Coasting out of Africa. Nature 405:24–25, 27

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L (1999) Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last ice age. Am J Hum Genet 65:1718–1724

Sutovsky P, Van Leyen K, McCauley T, Day BN, Sutovsky M (2004) Degradation of paternal mitochondria after fertilization: implications for heteroplasmy, assisted reproductive technologies and mtDNA inheritance. Reprod Biomed Online 8:24–33.

Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M (1995) The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. Am J Hum Genet 57:1463–1475

Tambets K (2004) Towards the understanding of post-glacial spread of human mitchocondrial DNA haplogroups in Europe and beyond: a phylogenetic approach. PhD, University of Tartu, Tartu

Tanaka M, Cabrera VM, Gonzalez AM, Larruga JM, Takeyasu T, Fuku N, Guo L-J, et al. (2004) Mitochondrial genome variation in eastern Asia and the peopling of Japan. Genome Res 14:1832–1850

Taylor RW, McDonnell MT, Blakely EL, Chinnery PF, Taylor GA, Howell N, Zeviani M, Briem E, Carrara F, Turnbull DM (2003) Genotypes from patients indicate no paternal mitochondrial DNA contribution. Ann Neurol 54:521–524

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

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 (2005b) 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

Thorburn DR, Dahl HH (2001) Mitochondrial disorders: genetics, counseling, prenatal diagnosis and reproductive options. Am J Med Genet 106:102–114.

Thorne A, Grun R, Mortimer G, Spooner NA, Simpson JJ, McCulloch M, Taylor L, Curnoe D (1999) Australia's oldest human remains: age of the Lake Mungo 3 skeleton. J Hum Evol 36:591–612

Torroni A, Bandelt H-J, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus ML, Bonne-Tamir B, Scozzari R (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62:1137–1152

Torroni A, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994) Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America. Proc Natl Acad Sci USA 91:1158–1162

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC (1993a) Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590

Torroni A, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH, Comuzzie AG, Wallace DC (1993b) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet 53:591–608

Turney CSM, Bird MI, Fifield LK, Roberts RG, Smith M, Dortch CE, Grun R, Lawson E, Ayliffe LK, Miller GH (2001) Early human occupation at Devil's Lair, southwestern Australia 50,000 Years Ago. Quaternary Research 55:3–13

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley R, Oefner PJ, Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43–62

van Andel T, Davies W, Weninger B (2003) The human presence in Europe during the last glacial period I: Human migrations and the changing climate. In: van Andel T, Davies W (eds) Neanderthals and modern humans in the European landscape during the last glaciation. McDonald Institute for Archaeological Research, Cambridge, pp 31–56

Vasil'ev SA, Kuzmin YV, Orlova LA, Dementiev VN (2002) Radiocarbon-based chronology of the Palaeolithic in Siberia and its relevance to the peopling of the New World. RADIOCARBON 44:503–530

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, et al. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA 98:10244–10249

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

Wiuf C (2001) Recombination in human mitochondrial DNA? Genetics 159:749–756

Wooding S, Ostler C, Prasad BV, Watkins WS, Sung S, Bamshad M, Jorde LB (2004) Directional migration in the Hindu castes: inferences from mitochondrial, autosomal and Y-chromosomal data. Hum Genet 115:221–229

Yao Y-G, Kong Q-P, Bandelt H-J, Kivisild T, Zhang Y-P (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635–651

Zhao X, Li N, Guo W, Hu X, Liu Z, Gong G, Wang A, Feng J, Wu C (2004) Further evidence for paternal inheritance of mitochondrial DNA in the sheep (Ovis aries). Heredity 93:399–403

Zsurka G, Kraytsberg Y, Kudina T, Kornblum C, Elger CE, Khrapko K, Kunz WS (2005) Recombination of mitochondrial DNA in skeletal muscle of individuals with multiple mitochondrial DNA heteroplasmy. Nat Genet 37:873–877

KOKKUVÕTE

India eelajaloo käsitlus mtDNA vaatevinklis

1. Tänaste India kasti- ja hõimurahvaste juured ulatuvad ühisesse eellas-populatsiooni mis asustas Lõuna-Aasia Ülem-Pleistotseenis.
2. Kõik India erinevates populatsioonides, sealhulgas hõimurahvaste seas esinevad mtDNA variandid kuuluvad ühte kahest Euraasia asutaja-haplo-grupist – haplogruppi M või N. Samas võivad konkreetsete alamhaplo-gruppide sagedused populatsiooniti, eriti just triivitundlike hõimurahvaste puhul, oluliselt erineda.
3. Vastne alamhaplogrupp M2, mis on määratud mtDNA kodeeriva ala asen-dustega positsioonides 447G, 1780 ja 8502, on Indias levinud haplogrupi M alamhaplogruppidest nii kõige vanem (73,000±22,900 aastat) kui ka kõige sagedasem (10%).
4. 15% India emaliinidest kuuluvad kolme Lõuna Aasiale ainuomasesse haplogruppi M2, U2 (harud U2a, U2b ja U2c) ning R5. Nende haplo-gruppide koalestsentsiajad ulatuvad 50 kuni 70 tuhande aasta sügavusse – seega on nad Indias levinud haplogruppidest vanimad ja nende levikut võib seostada India algse asustamisega.
5. Täiustasime haplogruppide M ja R sisestruktuuri uute, Indiale ainuomaste alamhaplogruppide eristamisega ning uurisime nende fülogeograafilist levikut.
6. Tänasest India emaliinide geenitiigist pärineb ainult väike osa Holotseenis Lääne Euraasiast toimunud sisserännetest. See seab kahtluse alla laialt levi-nud hüpoteeside, mille kohaselt toimusid Indiasse sel perioodil ulatuslikud sisseränded, näiteks indo-aarialaste invasioon, paikapidavuse. Kuivõrd geenid ja kultuur võivad, kuid ei pruugi koos rännata, ei saa me ümber lükata väiteid, mille kohaselt kõik Indias täna räägitavad keeled on pärit väljastpoolt Indiat. Samas on draviidi keelte kohalik teke siiski tõenäolisem.
7. Leidsime, et haplogrupi M levikupiir läänes on India (60%) ja Iraani (5%) vahel. Sama terav piir lõikab Indiale omaste haplogrupi M alamhaplo-gruppide leviku Indiast itta ja põhja. Seega on kõige tõenäolisem, et ema-liinide geenitiik, millele pandi alus India esmase asustamisega, ei ole täna-seks asendunud, vaid arenenud kohapeal. Hilisemates demograafilise ajaloo peatükkides on sellele tiigile nii idast kui läänest lisatud vaid pisut garneeringut.
8. Haplogruppide M ja N autohtoonsete juurliinide olemasolu Indias ja Indiast idas ning haplogrupi M puudumine Indiast läänes on kooskõlas Euraasia asustamise Lõuna Rannikutee hüpoteesiga.

ACKNOWLEDGEMENTS

My deepest gratitude goes to my supervisors Prof. Richard Villems and Prof. Toomas Kivisild. Their guidance has been invaluable on my journey through the hills and valleys of the mtDNA world.

This work would have been impossible without the input of all the co-authors of the papers. I am grateful to all the people whom I have never met, but who have donated their "three drops of blood" for making the study possible. I am particularly thankful to Prof. Surinder S. Papiha for his introduction of us to Indian population genetics and him, as well as Dr. S. Mastana, collecting most of the Indian samples used throughout of this study. Also, this work would have been impossible without expert technical assistance by Ille Hilpus and Jaan Lind.

I bow to all my fellow students and researchers in the lab. I truly enjoy the fascinatingly international atmosphere in the lab where quite often you don't realise what language you currently use. Thanks to you all. And, Phillip, thanks for the new darts!

I thank my first supervisor, Prof Ülo Mander, for making it clear to me that science is fun.

I owe to my friends for vigorous discussions on matters of fundamental importance☺.

I would like to give special thanks to my parents, who didn't push me on this road, but probably saw the inevitable. You have been supportive all the way.

Most importantly, I thank my wife, Pille, for everything. Without you I'd be lost. Science is a fascinating thing but it pales next to a smile of a child. I hug Uku Kaarel and Ulla Marie and love you both for the smile.

PUBLICATIONS

Kivisild T, Bamshad M, Kaldma K, **Metspalu M,** Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R (1999). Deep common ancestry of Indian and western Eurasian mtDNA lineages. Current Biology 9: 1331–1334.

Kivisild T, Rootsi S, **Metspalu M**, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk H-V, Stepanov V, Gölge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R (2003) *The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations.* American Journal of Human Genetics 72:313–332.

Kivisild T, Rootsi S, **Metspalu M,** Metspalu E, Parik J, Kaldma K, Usanga E, Mastana S, Papiha S, Villems R (2003) *The Genetics of Language and Farming Spread in India*. In: Bellwood P, Renfrew C (eds) Examining the farming/language dispersal hypothesis. The McDonald Institute for Archaeological Research, Cambridge, pp 215–222

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, 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 Genetics 5:26

CURRICULUM VITAE

Mait Metspalu

Citizenship:	Estonian
Date and place of birth:	August 24th, 1975, Tartu, Estonia
Family status:	married, two children
Address:	Institute of Molecular and Cell Biology, University of Tartu and Estonian Biocentre, 23 Riia Street, 51010, Tartu, Estonia
Telephone:	+372 7 375 051
Fax:	+372 7 420 286
E-mail:	mait@ebc.ee

Education and professional employment

1982–1993	Tartu Miina Härma Gymnasium
1993–1994	Konrad Mägi Painting studio
1994–1999	BA in landscape ecology and physical geography, Tartu University. "Symbiotic nitrogen fixation in Alnus-Frankia symbiosis". supervisors: Prof. Ülo Mander and sen. sci. Krista Lõhmus. The work was awarded a prize by the Estonian Academy of Sciences.
1999–2001	MSc in molecular biology and evolution, Tartu University. "Common maternal legacy of Indian caste and tribal populations" supervisors: Dr Toomas Kivisild, Prof Richard Villems.
2001–present	PhD student in molecular biology and evolution, Tartu University
From 2005	Estonian Biocentre, scientist

Scientific work

Starting from 1998 I have been studying and working in the Department of Evolutionary Biology of the University of Tartu under the supervision of Prof. Richard Villems and Prof. Toomas Kivisild. I have been involved in projects, associated with the analysis of human mtDNA and Y chromosome variation in different (South) Asian populations. The results of studies are published in the following articles and book chapters:

Kivisild T, Bamshad MJ, Kaldma K, **Metspalu M**, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, 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 SS, 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 SS, Deka R, Chakraborty R (eds) Genomic Diversity. Kluwer Academic/Plenum Publishers, pp 135–152

Kivisild T, Papiha SS, Rootsi S, Parik J, Kaldma K, Reidla M, Laos S, **Metspalu M**, Pielberg G, Adojaan M, Metspalu E, Mastana SS, Wang Y, Gölge M, Demirtas H, Schnekenberg E, Stefano GF, Geberhiwot T, Claustres M, Villems R (2000) An Indian Ancestry: a key for understanding human diversity in Europe and beyond. In: Renfrew C, Boyle K (eds) Archaeogenetics: DNA and the population prehistory of Europe. McDonald Institute for Archaeological Research University of Cambridge, Cambridge, pp 267–279

Kivisild T, Rootsi S, **Metspalu M**, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk H-V, Stepanov V, Gölge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, 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 (eds) Examining the farming/language dispersal hypothesis. The McDonald Institute for Archaeological Research, Cambridge, pp 215–222

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, 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 Genetics 5:26

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

Metspalu M, Kivisild T, Bandelt H-J, Richards M, Villems R (in press) The pioneer settlement of modern humans in Asia. In: Bandelt H-J, Macaulay V, Richards M (eds) Human mitochondrial DNA and the evolution of Homo sapiens. Springer-Verlag, Heidelberg

Endicott P, **Metspalu M**, Kivisild T (in press) Genetic evidence on modern human dispersals in South Asia: Y chromosome and mitochondrial DNA perspectives. In: Petraglia M-D (ed) The Evolution and Diversity of Humans in South Asia. Springer/Kluwer Academic Publishers

in Estonian:

Metspalu M ja Tambets K (2004) Aasia asustamine kaasaegse inimese poolt: tulvavesi piki India Ookeani rannikut. Publicationes Instituti Geographici Universitatis Tartuensis 89, Toim. Ü. Mander, K. Remm, Tartu

Metspalu M, (2003) Eestlaste ema- ja isaliinidest Lääne-Euraasia kontekstis. LOODUSKAITSEALASEID TÖID VII, Toim. J. Eilart, J. Pärn T. Nuum, E. Kaur. Tartu Üliõpilaste Looduskaitsering, Tartu, 2003

Metspalu M, (2002) Mediaanvõrgud fülogeneesi kujutamisel. Võrkude teooria. Schola Biotheoretica XXVIII. Toim. I. Puura, T. Teder. Tartu: Sulemees, 2002.

CURRICULUM VITAE

Mait Metspalu

Kodakondsus:	Eesti
Sünniaeg ja -koht:	24 august, 1975, Tartu, Eesti
Perekonnaseis:	abielus, peres kaks last
Aadress:	Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituut, Eesti Biokeskus, Riia mnt.23, 51010, Tartu, Eesti
Telefon:	+372 7 375 051
Faks:	+372 7 420 286
E-post:	mait@ebc.ee

Haridus ja erialane teenistuskäik

1982–1993	Tartu Miina Härma Gümnaasium
1993–1994	Konrad Mäe nim Maalistuudio
1994–1999	Tartu Ülikooli bakalaureuseõpe loodusgeograafia ja maastiku-ökoloogia erialal. Bakalaureusetöö ”Sümbiontne lämmastiku fikseerimine Alnus-Frankia sümbioosis”. Juhendajad: Prof. Ülo Mander ja van. tead. Krista Lõhmus. Töö pälvis Eesti Teaduste Akadeemia preemia.
1999–2001	Tartu Ülikooli magistrantuur molekulaarbioloogia ja evolutsiooni erialal. Magistritöö “Common maternal legacy of Indian caste and tribal populations” Juhendajad: Dr Toomas Kivisild, Prof Richard Villems.
2001–...	Tartu Ülikooli doktorantuur molekulaarbioloogia ja evolutsiooni erialal.
2005–...	Eesti Biokeskus, teadur.

Teadustegevus

Alates 1998 a. olen õppinud ja töötanud evolutsioonilise bioloogia õppetooli juures. Olen olnud seotud mitmesuguste teaduslike projektidega, mille eesmärgiks on inimese mitokondriaalse DNA ja Y kromosoomi varieeruvuse uurimine erinevates (Lõuna) Aasia populatsioonides. Saadud tulemused on publitseeritud lk. 133 loetletud artiklites.