

# Peopling of South Asia: investigating the caste–tribe continuum in India

Gyaneshwer Chaubey,<sup>1\*</sup> Mait Metspalu,<sup>1</sup> Toomas Kivisild,<sup>1,2</sup> and Richard Villems<sup>1</sup>

## Summary

In recent years, mtDNA and Y chromosome studies involving human populations from South Asia and the rest of the world have revealed new insights about the peopling of the world by anatomically modern humans during the late Pleistocene, some 40,000–60,000 years ago, over the southern coastal route from Africa. Molecular studies and archaeological record are both largely consistent with autochthonous differentiation of the genetic structure of the caste and tribal populations in South Asia. High level of endogamy created by numerous social boundaries within and between castes and tribes, along with the influence of several evolutionary forces such as genetic drift, fragmentation and long-term isolation, has kept the Indian populations diverse and distant from each other as well as from other continental populations. This review attempts to summarize recent genetic studies on Indian caste and tribal populations with the focus on the information embedded in the socially defined structure of Indian populations. *BioEssays* 29:91–100, 2007. © 2006 Wiley Periodicals, Inc.

## Introduction

South Asia refers to the southern geographic region of the Asian continent that is surrounded by Iranian Plateau from the west, Himalayas from the north and east and by the Indian Ocean from the south. The Indian subcontinent, often used as a synonym for South Asia, is drawn presently within the political borders of India, Pakistan and Bangladesh. Throughout its history South Asia has been a mixing pot of ethnic groups, languages and cultures. There is a fundamental resemblance in various rituals practiced by people in different regions and, therefore, shared ritual patterns can account for

some unity in the varieties of the religious beliefs that we can see in South Asia over a long time. The written phase of history, covering a mere two millennia, includes numerous accounts of invasions to the subcontinent and gives evidence of a multitude of cultural contacts of India with its close and more distant neighbors. There are reasons to believe that such contacts and interactions extend to a far longer prehistoric period of the region. Therefore, it is not surprising to find Indian population genetically and socio-culturally so highly diverse today.

Though archaeological evidence for the existence and activities of the hominid populations is profuse throughout the Indian subcontinent during the middle and late Pleistocene, the fossil record of hominids during this period is rather poor. A partial hominid cranium (around 250,000–300,000 years old) has been found in the Narmada River at Hathnora in Madhya Pradesh.<sup>(1)</sup> This cranium is currently attributed to *Homo heidelbergensis*.<sup>(2)</sup> The earliest fossils of modern humans in South Asia have been unearthed in Sri Lanka and are dated to around 28–34 thousand years before present.<sup>(3,4)</sup> The earliest city to be discovered in India was Harappa in Punjab (presently in Pakistan). Further down in the Indus valley another ancient city was excavated and this was Mohenjo Daro in Sindh.<sup>(5)</sup> The archaeologists refer to the civilization of these ancient cities as Harappa or the Indus Valley Civilization, because both of these sites and other sites sharing the same culture were found in the Indus valley.

Fluctuating environmental conditions have often had a global impact on the prehistoric continuity and demographic succession of populations around the world, including South Asia.<sup>(6,7)</sup> The impact of environment not only would have affected the spread of cultural information within the region but also would have determined the number of possible dispersals into the region.<sup>(7)</sup> One such event that had a dramatic influence on climate in the areas around the Indian Ocean was the Mt. Toba supereruption in Sumatra approximately 74 thousand years ago.<sup>(8)</sup> The archaeological record is not conclusive about the existence of modern humans in Asia before the Toba eruption. However, it has been hypothesized that modern humans passed through a genetic bottleneck roughly around this time frame with only a few tens of thousands of survivors whose numbers have expanded thereafter in different

<sup>1</sup>Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu and Estonian Biocentre, Tartu, Estonia.

<sup>2</sup>Leverhulme Centre of Human Evolutionary Studies, The Henry Wellcome Building, University of Cambridge, UK.

Funding agency: This work was supported by Estonian Basic Research grant SF0182474 (to RV) and Estonian Science Foundation grant 5574 (to TK).

\*Correspondence to: Gyaneshwer Chaubey, Department of Evolutionary Biology, Tartu University and Estonian Biocentre, Riia 23, Tartu, 51010, Estonia. E-mail: gyanc@ebc.ee

DOI 10.1002/bies.20525

Published online in Wiley InterScience (www.interscience.wiley.com).

continents, leaving behind specific signatures in the pairwise mismatch distributions of their mitochondrial DNA sequences.<sup>(9–11)</sup> This view, however, relies on the assumption that the out-of-Africa migration had already occurred before the Toba eruption. Given the evidence of ash layers in India, it is possible that the human populations in South Asia may have become largely extinct and that the subcontinent was subsequently re-colonized by a migration from the east.<sup>(12)</sup> An alternative view, based on the coalescent analysis of complete mitochondrial DNA sequences (see Box 1), suggests that the out-of-Africa migration was launched after the Toba eruption and that South Asia may have been one of the first corners of Eurasia where modern human populations started to expand and diversify genetically (reviewed in more detail in Richards et al.<sup>(13)</sup>). Geographic expansion of modern humans outside Africa must have resulted in a significant increase in human population size. Why and technically how modern

### Box 1.

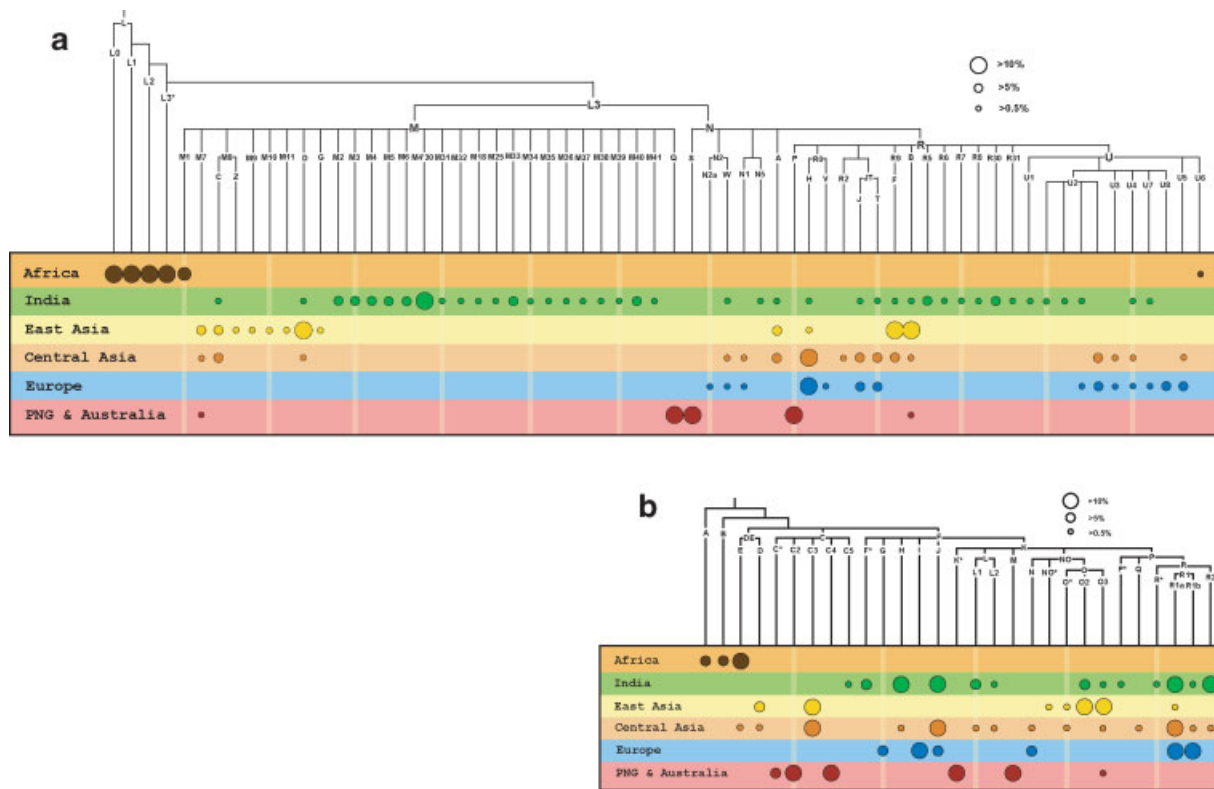
Coalescence approach. Coalescence is a central concept in evolutionary genetics that stands for the reduction of the number of common ancestors of the extant genetic lineages when looking at time in retrospect. While genetic sequences accumulate mutations over time and diverge from each other, the coalescence methods attempt to reconstruct the process of divergence by assuming that sequences most similar to each other share a more recent common ancestor than those that are more distant. Most commonly used methods for inferring coalescent trees involve maximum parsimony, maximum likelihood, neighbour-joining, and Bayesian methods.<sup>(88)</sup> Knowing the rate at which mutations occur allows the coalescent methods to assign probabilistic age estimates to the once living ancestors, often with wide error margins, though. The evolutionary rate of human mtDNA sequences, calibrated by fossil dates of human and chimpanzee divergence, assumes, for example, that one synonymous transition in protein coding genes arises in a matrilineage (an array of mother's mother's ...mother) approximately every seven thousand years.<sup>(89)</sup> Thus, a sample of sequences that show, on average, a distance of three mutations to their most recent common ancestral sequence, is dated as 20 thousand year old with certain confidence intervals which depend on the number of the individuals sampled and the structure of the tree. Three ancestral sequences M, N, and R, dated to approximately 40–70 thousand years old,<sup>(44,89)</sup> capture virtually all non-African mitochondrial sequences and are suggested thus to be the founding ancestral lineages involved in the settlement of Eurasia and Oceania.

humans practising a hunter–gatherer lifestyle managed to colonize Eurasia and Australia in late Pleistocene is an important question relevant to the very definition of modern humans.<sup>(7,14)</sup>

Several studies have pinpointed the role of genetic structure in the Indian subcontinent in understanding the initial settlement of Eurasia by modern humans.<sup>(15–24)</sup> Differences in allele frequencies in and between human populations have been the major platform for studying the history and structure of these different populations. A comprehensive analysis of classical genetic polymorphisms involving more than 100 genes, showed South Asian populations lying in between the populations of Southeast and West Asia.<sup>(25)</sup>

Nonetheless, from the individual maps of single genes the genetic landscape of India turned out to be particularly complex in gene-to-gene comparisons suggesting that the peculiar tribal–caste social structure has generated “a multitude of endogamous pockets”. Thus, an unusually high level of genetic differentiation can be observed in South Asia over short geographic distances. A world-wide study of *Alu* and microsatellite polymorphisms revealed the certainty of at least 90% for 60 markers, and 99%–100% with 100 loci, serving to classify by genetic information individuals as belonging either to African, European or East Asian continental clusters.<sup>(26)</sup> However, when individuals of South Asian ancestry were added to this analysis, they failed to form a discrete structural unit like the others. Another study, using 45 autosomal STR loci, revealed that Indian populations form a number of distinct genetic clusters that may be as dissimilar to each other as European populations are from African, for example. Nevertheless, the tribal and caste populations were, on average, genetically closer to each other than to any other continental group.<sup>(27)</sup> It is evident, therefore, that the high level of heterogeneity in Indian populations, governed by high level of endogamy created by numerous social boundaries, along with the influence of general evolutionary forces such as genetic drift due to long-term isolation, has kept the Indian gene pool distinct from other continental populations.

Recently, mtDNA and Y chromosomal genetic studies have provided a substantial contribution to the understanding of human origins and dispersal patterns. Mitochondrial DNA surveys on populations worldwide have divulged continent-specific distributions of basic mtDNA clades or haplogroups.<sup>(28–31)</sup> Phylogenetic reconstructions further reveal that the extant mtDNA variation outside Africa arises from just two basic founder lineages (designated M and N), which are two of many daughter lineages of an African-specific haplogroup L3 (Fig. 1a).<sup>(18,23,32–36)</sup> Besides mtDNA evidence, the out-of-Africa scenario has found support from various sources of genetic and archaeological data (for recent reviews see Refs 37,38). In the debate about single versus multiple dispersals of anatomically modern humans out of Africa, recent mtDNA and Y chromosomal studies have produced evidence supporting a single early



**Figure 1.** The summary and distribution of Global mtDNA (a) and Y chromosome (b) topology. The haplogroup nomenclature and data follows previous studies<sup>(15,16,18,21,22,24,32,33,36,40,41,45,50–52,56,61,64,66–68,85,92,93)</sup> but is updated here according to our unpublished data. The lower panel (b) shows the Y chromosomal clades and their frequency in Eurasia, where subclades F\*, H, L1 and R2 are autochthonous to India and outside India are found at low frequencies only in Central Asia. The upper panel (a) depicts the mtDNA clades and their frequency in Eurasia. African-specific branches of the tree are simplified to show only the very basic lineages and identify the root in L3 for the pan-Eurasian founder lineages M and N (including R). The region specificity of different haplogroups is shown in color as indicated. Native American mtDNA variation is a subset of East Asian variation and is not specified in the figure. There is a multitude of haplogroup M subclades that are autochthonous to India but, most importantly, all the pan-Eurasian founder lineages—M, N and arising from the latter, R—sport deep-rooting subclades that are found only in South Asia. This is the strongest argument pointing to the pivotal role South Asia must have had during the initial out-of-Africa exodus some 40 to 60 thousand years ago. Note that haplogroup U2 has one West Eurasian-specific subclade U2e and that haplogroups U7, W and R2 all have South Asian-specific subclades (not shown on the figure) with coalescent times well beyond the Last Glacial Maximum.

migration that brought ancestral mitochondrial (M, N and deriving from the latter, R) and Y chromosomal (C, D and F) lineages to Eurasia and Oceania, suggesting their co-migration along the southern route in one wave. This view has also gained support from archaeological evidence.<sup>(39)</sup> The distribution of basic mtDNA founders (M and N) is not uniform globally. Most western Eurasians carry mtDNAs that belong to clades derived from founder haplogroups N and R,<sup>(31,40,41)</sup> whereas N (including R as its daughter group) and M are equally present throughout Asia, Melanesia and Australia (Fig. 1a).<sup>(15–18,23,35,36,41–46)</sup> In this review, we focus on the correlations between genetic variation, the languages, and geographic spread of tribes and low status (scheduled) castes in India. We also shed light on the reconstruction of prehistory of Indian gene pool by haploid genome marker studies.

### The ongoing debate...

The origin of the Indian caste system and the Indo-Aryan speaking populations is a matter of intense academic debate with its history going back to Sir William Jones, who, as a judge in the Presidency of Bengal, originally reported striking similarities between Sanskrit, Greek and Latin at the end of the 18<sup>th</sup> century. He suggested a common source for all these languages, while Max Müller, another distinguished orientalist, added, a half century later, the argument that the Aryans might have migrated to India around 1500 BC. More than two centuries of extensive interdisciplinary scholarship, in particular after the 20<sup>th</sup> century discovery of the Indus Valley Civilization and the establishment of the main facts about the neolithization of South Asia, has greatly widened and enriched the panorama of the understanding of the peopling of South

Asia in the global context. The study of classical markers put South Asia in between the populations from West and South-east Asia.<sup>(25)</sup> Furthermore, the 1st principal component, in an analysis of 69 genes from 42 populations of Asia, explains 17.7% of the variation from southeast and East Asians into India.<sup>(25)</sup>

More recently, several genetic studies have added support to this theory.<sup>(47–51)</sup> Certain genetic variants were found to be shared among Indian and European populations. However, subsequent studies using more representative sample sizes and, importantly, a higher level of molecular resolution, have established that, even though Indian and West Eurasian populations share a common genetic ancestry in late Pleistocene, gene flow into India during the period of the proposed Aryan invasion has been minimal.<sup>(15,17,18,52)</sup> As yet the evidence is equivocal and there is no genetic signal for a major genetic component associated either with the spread of Indo-Aryan languages or the caste system within India.<sup>(53)</sup>

The complex structure of the Indian caste system and its origin is another matter of dispute. The caste society definitely plays, and may well long have played, an essential role in the genetic and social structuring of South Asian populations.<sup>(54)</sup> The haploid genetic study of castes was, perhaps, started by a chain of papers relying on the idea of the male-mediated Indo-Aryan invasion, which supposedly pushed the indigenous Dravidian populations southwards, and established the Aryans at the top of the caste hierarchy.<sup>(21,55,56)</sup> These studies suggested that the caste and tribal groups differ significantly in haplogroup frequencies. Moreover, caste groups were found more homogeneous for Y chromosome variation and more closely related to Central Asian groups than to Indian tribal or any other Eurasian groups.<sup>(56)</sup> These results might, however, have been affected by limited sampling and depth of analysis of Indian tribal and caste groups because other extended studies have failed to confirm the general tribal distinction from the castes.<sup>(22)</sup> One important aspect of the Indian caste system, which has a substantial impact on the inferences that one can make from the caste/tribal genetic variation, is that the definition of the castes has been historically fluid. Specifically, the scheduled tribes have been gradually incorporated into the caste system as scheduled castes.<sup>(57)</sup>

Yet, another question concerns the initial settlers of India. There are several studies which state that the Austro-Asiatic tribal groups represent the first settlers of India most closely.<sup>(21,58,59)</sup> Basu et al.<sup>(21)</sup> suggested that haplogroup M2, which is one of the oldest Indian-specific mtDNA haplogroups,<sup>(18)</sup> is particularly frequent and diverse among Austro-Asiatic speaking tribes of India. This argument was, however, based on an inaccurate assignment of the tribal samples into the M2 haplogroup on the basis of hypervariable segment I (HVS I) motif 16223–16319, which occurs independently in many different haplogroups. Moreover, because haplogroup M2 is spread across the borders of Indo-European, Dravidian and Austro-

Asiatic language families and is shared by both tribal and caste communities,<sup>(52)</sup> it is problematic to relate the initial Palaeolithic settlers of India on the basis of certain linguistic affiliations.

### Haploid genetic view on South Asia

#### MtDNA studies

The South Asian mtDNA pool is largely made up of basal autochthonous deep rooting lineages arising from the three pan-Eurasian founder types M, N and R, the latter nested in N (Fig. 1a). These Indian-specific mtDNA clades are virtually absent from the adjacent European and East Asian mtDNA pools. The autochthonous mtDNA haplogroups in Indian populations include: U2a,b,c, R5–8, R30, R31, N1d and N5 in haplogroup N<sup>(35)</sup> and M2–6, M30–47 in haplogroup M (Fig. 1a) (see Refs 24,38 and our unpublished data). Almost 60% of Indians have their maternal root in haplogroup M.<sup>(15,18)</sup> The pattern of mtDNA haplogroup distribution does not follow the caste/tribal or language group-mediated structuring of the extant South Asian population.<sup>(15,16,24,52)</sup>

The phylogenetic structure (see Fig. 1a) of major M sub-clusters in the Indian subcontinent suggests that the region was settled soon after the African exodus<sup>(15–17,23,24)</sup> and that there has been no extinction or replacement of the initial genetic footprints. Overall, the Eastern Eurasian-specific haplogroups viz A, B, C, D, E and G are found at low frequencies while only their share in the northeastern part of India rises to one half or more.<sup>(52,60)</sup> Haplogroup U is common in West Eurasian populations while its three sub-clades U2a,b and c are present in South Asia and share a deep-rooted coalescence with European founding populations.<sup>(15)</sup> The pool of mtDNA lineages found in India is partially amalgamated with eastern and western Eurasian mtDNA haplogroups (Fig. 1a) of both ancient and young MRCA (most recent common ancestor) age. Geographically the zone of admixture of West and South Asian maternal lineages is concentrated towards north-western India.<sup>(52,61)</sup> The major body of the shared haplogroups consists of basal deep-rooting branches of haplogroup R. This link is likely to be as ancient as the out-of-Africa migration and does not contribute to the (state of) “West and South Asian recent admixture” The minor share of lineages nested in haplogroups, such as in TJ (1.7%), U5 (0.20%), U4 (0.16%), and K (1.3%) likely relate primarily to migrations during the Holocene period, while the exact source and timing of such migrations (either due to Neolithic Farmers, nomadic Bronze Age people, or any later migrations, including Huns, Moghuls, Greeks, etc.), is difficult to establish.<sup>(15–18,51,62–64)</sup> Moreover, the Islamization of India during the Muslim dynasty in the 17<sup>th</sup> centuries did not introduce significant genetic contribution to the Indian gene pool.<sup>(21,65)</sup> The internal structure of haplogroup M and N (N1d and N5, R5–R8, R30, R31, U2a,b,c) lineages in India, as revealed now by the analysis of complete mtDNA genomes, reflects their basically autochthonous development (Fig. 1a).<sup>(24,35,36)</sup>

### Y chromosomal studies

The uniparentally inherited non-recombining haploid Y chromosome is a widely used marker for assessing the origins of populations along the paternal descent line.<sup>(66)</sup> Most Indian communities trace their origin back along the male 'gothra' or clan, which is often the basis of endogamous marriage networks. It is notable that the gothra system exists in caste as well as in tribal populations. The majority of Y gene pool of South Asia contains haplogroups C, H, J, R1a, R2, L, and O2a (Fig. 1b).<sup>(18,22)</sup> The high STR variance and widespread nature in Indian subcontinent of haplogroups C5, F\*, H, R2 and L1 has usually been considered of indicative to their indigenous origins in the subcontinent.<sup>(64)</sup> A few studies have suggested haplogroup R1a, with its wide geographic spread including Eastern Europe and Central Asia, as a potential marker of the Indo-Aryan invasion that introduced the caste system to India, as the frequency of this haplogroup was found to be specifically higher among the caste groups.<sup>(21,50,58)</sup> Several other papers, however, have argued against such a simple, essentially single alpha-male lineage initiated migration scenario, which receives no significant support from the maternally-inherited gene tree.<sup>(22,64)</sup> The higher variance of STRs in the Indian R1a lineages as compared to those from Central Asia further weakens such a scenario, implying a strong founder effect.<sup>(18)</sup> However, the current lack of sufficient SNP marker resolution makes it difficult to infer the geographic origin of haplogroup R1a. The high frequency and STR diversity of haplogroup R2 in Indians corroborates its Indian origin.<sup>(18,22,64)</sup> It has also been reported in Iran and Central Asia<sup>(50,67)</sup> with marginal frequency, which more likely suggests a recent migration from India. It is present at high frequency (53%) among Gypsies of Uzbekistan, known to have historically migrated out from India.<sup>(50)</sup> Interestingly, this haplogroup is absent or infrequent among Gypsies of Europe whose predominant Y chromosome haplogroup is H.<sup>(68)</sup> Haplogroup O2 spread is characteristic mostly of the Austro-Asiatic speaking populations of India and South East Asia. The predominance of the O haplogroup and its sublineages in populations of Eastern and Central Indian suggest a SE Asian origin of Indian Austro-Asiatic and Tibeto-Burman speakers, with the latter being likely very recent immigrants.<sup>(63)</sup>

### Caste-tribe continuum

The subject of the origins of the caste system (see Box 2) is one of the most interesting as it deals with a social organization that has been the major distinguishing feature of the civilization of South Asia. Only about 8% of the Indian population today has a tribal affiliation (2001 census). However, at least during the last 500 years or more, the definition of caste/tribe has constantly been changing, due to the influence of the ruling authorities. Such regulatory reappraisals should not be seen as arbitrary political acts of these authorities but as reflecting various external changes, including the demographic context of the

#### Box 2.

Scheduled castes and scheduled tribes. Scheduled Castes and Scheduled Tribes are communities in India that are given a special status by the Constitution of India. Scheduled castes were considered as 'Sudra' in the classical 'Chaturvarna' caste system while Scheduled tribes were considered 'outcastes' and were not a part of the Indian caste system. 'Chaturvarna' is a division of castes into four categories: Brahmin, Kshatriya, Vaishya and Sudra, which represent the social structuring of Hindu caste system society, which is highly stratified even further within each of these basic categories. Amongst the Sudra, the tribal population represents the most disadvantaged (often officially termed as "backward") group in the highly ripped Indian society. The caste and tribal people of the lowest hierarchical status have during the last century been enlisted by Indian Government and given a legal classification as "Scheduled" castes and tribes. Several specific provisions have been made in the Indian constitution for raising the socio-economical status of these "Scheduled" groups.

habitat, available to different subpopulations. Crop cultivation resulted in the loss of the traditional habitat of hunter-gatherers by deforestation, fragmenting and marginalizing numerous such populations, many of whom were assimilated into agriculturally based subsistence economies, thereby catalyzing the change of group membership from scheduled tribe to scheduled caste. Some of them began pursuing criminal practices for their livelihood and became classified formally as criminal tribes (see Box 3). This dynamic process of absorption of the tribal populations into the caste system can explain the contrasting geographical distributions of the Scheduled Tribes and the Scheduled Castes communities. The scheduled castes make up 16% of the total Indian population (2001 census). The present distribution of Scheduled caste and Scheduled tribes in India (Fig. 2) clearly illustrates the contrasting scattering of caste and tribe affiliations. The present tribal population is restricted to the Western Ghats in Maharashtra, Karnataka and Kerala, the Eastern Ghats in Andhra Pradesh and Orissa, the Chota Nagpur plateau in Bihar and in West Bengal, entire Northeast India outside the Brahmaputra valley along with Aravalli hills in Rajasthan and Gujarat, the Vindhya and Satpura hills in Madhya Pradesh (Fig. 2b). In contrast, in the plains of Punjab, Haryana, Rajasthan, Uttar Pradesh, Bihar, West Bengal, Orissa, Madhya Pradesh, Karnataka, Andhra Pradesh, Tamil Nadu and Tripura, the percentage of Scheduled Tribes population is negligible and that of the Scheduled Castes is high (Fig. 2).

### Box 3.

Criminal tribes. The British government in the 19<sup>th</sup> century branded the communities indulging in crime as criminal tribes and established special settlements for them. After independence, this pejorative label was removed and they are now officially designated as denotified and nomadic tribes. Many of them continue to practice hunting and gathering and produce various craft items for the village people. The important communities in this category are Kanjars, Dharkars, Nut, Haburas, Bediyas, Bahelias, Bhantus in the Ganga plains; Sansis and Kalbeliyas in Rajasthan; Chharas in Gujarat; Pardhis and Kuchbandhis in Madhya Pradesh, Pardhis and Vaidus in Maharashtra, Lambadi in Andhra Pradesh and Kallars in Tamilnadu. The expansion of agriculture took a heavy toll of forests and wildlife. As a consequence, hunter-gatherers were compelled to accept agriculture and associated occupations and become assimilated into the steadily expanding caste-based rural and urban system. Hunter-gatherers, who were too conservative to adopt the economically beneficial yet arduous agricultural way of life, have persisted with their original lifestyle right up to the present. However, because of the steady reduction of their habitat and traditional food resources by continuous encroachment by rural and urban populations, they have been forced to adopt one or more of the additional occupations like providing various craft items produced from rope, grass and stone, while still sometimes resorting to theft. Such groups like the Kanjars, Sansis, Dharkars Nut, Lambadi and Pardhis, to name only a few, are found all over the country, including the outer areas of metropolitan cities.<sup>(90)</sup>

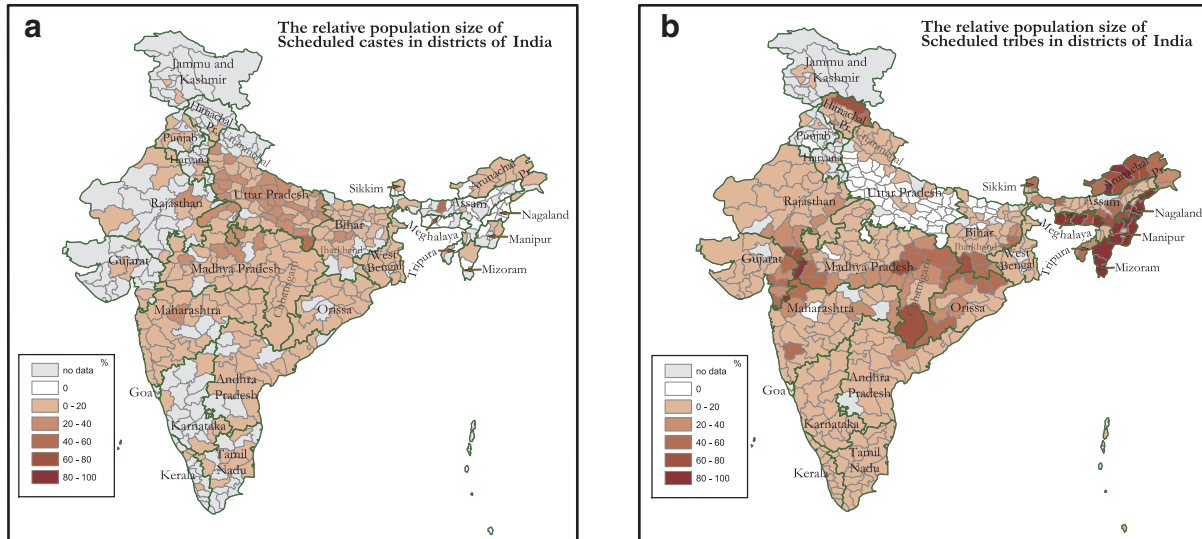
### Languages and genes

Linguistically, the different Indian populations belong to four major language families: Indo-European, Dravidian, Austro-Asiatic and Sino-Tibetan. The Indo-European group is the most widespread and is found in North, Central and Western parts of India. The geographic spread of the Indo-European language family is immense—stretching from Iceland to the Indian subcontinent. The origin of this language family is still unresolved. Some of the studies claim that Anatolia (what is today central and south-eastern Turkey) is the cradle from which the Indo-Europeans dispersed to their historical habitats.<sup>(69,70)</sup> This view presumes that the Hittites and other peoples who spoke related languages (Luvian, Pallaic) were autochthonous. Historians and archaeologists, however, disagree with this view and regard the Hittites as intrusive to Anatolia<sup>(71–73)</sup> while some others regard Saptasindhu in what is today Northwest India and Pakistan as the original homeland

of the Indo-Europeans.<sup>(74–76)</sup> Another hypothesis suggests its origin in the steppes of the Black Sea region<sup>(77)</sup> where the domestication of the horse and discovery of the wheel around 4000 BC facilitated the spread of the languages together with that of the people.<sup>(78,79)</sup>

Dravidian-speaking populations are restricted, with a few limited exceptions, to South India. It has been hypothesized that the proto-Elamo-Dravidian language spoken by the, now extinct, Elamites in southwestern Iran, spread eastwards with the movement of farmers from this region to the Indus Valley and the Indian subcontinent.<sup>(25,80,81)</sup> The Elamite relation of Dravidian language is debated in another study.<sup>(82)</sup> Further research on the spread of agriculture and associated vocabulary has supported an indigenous origin of Dravidians. Fuller<sup>(83)</sup> proposed the indigenous origin of Dravidic languages in South India. He discussed a new model for the origin of the Dravidian languages based on archaeobotanical evidence and suggested that they originated in South India. He argued for their dispersal in two directions; one towards Orissa and Bihar and another 'out-of-India' through Gujarat. Further, he argued independent origins of rice, millet and gram domesticated in the Ganges Valley and South India. The mtDNA studies find a closer affinity of Brahui (a Dravidian outlier population in Pakistan) with Indo-Iranian than with the Indian Dravidian.<sup>(61)</sup> A recent genetic study on the Y chromosome<sup>(64)</sup> also advocated this model but linking with the Y chromosomal haplogroup L1 does not fit with other South Indian genetic studies.

There are approximately 150 different Austro-Asiatic languages spoken in communities that are scattered from the Indian subcontinent to Vietnam in the east and the Malay Peninsula in the south. The wide spread range of this family is considered to be due to the agricultural expansion of populations whose homeland was in South China.<sup>(84)</sup> In India, the Austro-Asiatic speakers are dispersed mostly in the central and eastern parts of the country. A few tribes of Central India, Chotanagpur and Orrisa, speak a Mundari branch of this language while another branch Mon-Khmer speakers are Khasi in Meghalaya. The high frequency of Y chromosomal haplogroup O2a (M95) in Austro-Asiatic group<sup>(22,85)</sup> is consistent with their homeland in Southeast Asia. However, mtDNA studies have revealed different demographic histories for maternal lineages of the Mundari and Mon-Khmer groups.<sup>(86)</sup> While the Mundari groups carry predominantly Indian-specific mtDNA haplogroups, the Mon-Khmer speakers are affiliated with the East Asian mtDNA stock.<sup>(86)</sup> The Sino-Tibetan speakers are concentrated in the northeastern parts of the country. The Sino-Tibetan language is thought to have originated in China and spread from the Yellow river basin into Burma and the Himalayas.<sup>(87)</sup> Genetic studies on North Eastern tribes of India divulged a closer relationship to East Asians than to other Indians,<sup>(60)</sup> which is generally consistent with their recent spread to India from East Asia.<sup>(63)</sup>



**Figure 2.** District wise distribution of the scheduled castes (**a**) and scheduled tribes (**b**) in India. Substantial differences in the reported population of scheduled tribes in neighboring districts belonging to different states are best explained by administrative reasons. In southern Uttar Pradesh, the reported frequency of Scheduled Tribes is next to zero, whereas in all the districts of adjacent states that have common borders with Uttar Pradesh in the north the frequency of Scheduled Tribes is notable. Note, for example, the high percentage (54%) of scheduled tribe populations in Surguja, Chhattisgarh and nil in the adjacent Sonbhadra, Uttar Pradesh where, in contrast, the scheduled caste population is reported to be numerous (42%). The situation along the border between Madhya Pradesh and Maharashtra is rather similar in nature but opposite in numbers. Geography may be behind this pattern there because the forested hilly landscape of southern Madhya Pradesh would facilitate cultural isolation of the Scheduled Tribes dwelling in the region. This illustrates the blurred border between scheduled castes and tribes of India. Communities are assigned to either category by local administration which apparently demonstrates a state bias. Data from Census of India, 1991 accordingly, the administrative borders predate the reform in 2001. The new states are indicated by text only.

The extant mtDNA and Y-chromosome pools of all the language groups have been studied with the attempt to reveal whether any of these can be singled out as more “autochthonous” than others. However, as argued in Metspalu et al.<sup>(52)</sup> this would be highly problematic, first, because the language families involved are generally believed to be far younger than the time frame required for the peopling of India. Secondly, such “autochthonous” Indian-specific mtDNA and Y chromosome lineage groups are widely spread across language borders in the subcontinent, as exemplified by the Mushar group (see Box 4), and the putative language shifts make it hard to infer the original tongue for every population studied even during the historic period and perhaps impossible for earlier times. Thus, the present-day linguistic affinities of different Indian populations per se are perhaps among the most ambiguous and even potentially controversial lines of evidence in the reconstruction of prehistoric demographic processes in India.

## Conclusions

Recent progress in the understanding of topology of phylogenetic trees of the haploid mtDNA and Y-chromosomal

genomes combined with increasingly more detailed phylogeographic mapping of their corresponding branches globally and in the Indian subcontinent more specifically, conclusively show that the gene pool of South Asia is made up essentially from the same basic components as other non-African gene pools while the local differentiation and long-term genetic isolation have provided numerous diverse local genetic variants stemming out from these founders. Most of the Indian-specific mtDNA haplogroups show coalescent times 40,000–60,000 YBP. Their virtual absence elsewhere in the world suggests only a limited gene flow out from the subcontinent over a long time span, probably since the first *in situ* expansion phase in Late Pleistocene. Further efforts are needed to reveal evolutionary, temporal and spatial trajectories of some Y-chromosomal haplogroups, in particular that for haplogroup R, which is omnipresent in large continental areas of Eurasia. This is particularly important because of an immediately apparent substantial overlap between the spread of Indo-European languages and NRY haplogroup R derivatives.

In parallel, progress in molecular resolution of the mtDNA and Y-chromosomal genealogies allows identification of haplogroups, likely introduced to South Asia at later stages. A clear subset of mtDNA variants characteristic for East and

**Box 4.**

Language shift. Language shift is a cultural process in which an expanding population changes their language of that of a surrounding population with only a minor contribution of that population genes.<sup>(25)</sup> This process makes for discordance between languages and genes and that largely hampers the journey of gene with language. An important process that has shaped the present (spatial) distribution of languages in South Asia is language shift by indigenous populations. The Mushar community is one of the best examples of language shift in India. This community is dispersed mainly in Uttar Pradesh, Bihar, Jharkhand, Chhattisgarh and Madhya Pradesh and is known to have spoken the Mundari dialect of the Austro-Asiatic language family in the recent past.<sup>(91)</sup> But now they have by and large adopted Indo-European language from the surrounding populations. (Mushar literally means mouse eaters Hindi; *Mus*-‘Mouse’, *her*-‘eater’). Some of them are still hunters and gatherers and continue to speak the Mundari language. Y chromosome and mtDNA studies on this population and neighboring Indo-European populations revealed an almost impermeable genetic boundary between them. This suggests that the cultural process in India is one where different communities live side by side, interchanging ideas, foods and goods but not genes. In certain areas, however, the juxtaposed communities are more or less independent of one another, kept separate by the hierarchal category in the caste system.

Southeast Asia (Fig. 1a) are expectedly most frequent in northeastern provinces of India among populations speaking Tibeto-Burman languages. Yet the spectrum is more complex, revealing recent and, possibly, also earlier language shifts in different populations, as well as sex-specific admixture patterns.

Several open and complex questions, such as the origin of caste (gothra) system, the arrival of the major language families and their spread in South Asia, need to be solved by combining different disciplines like archaeology, historical linguistics and genetics. In particular, new promising high-throughput technologies to study autosomal genetics of South Asian populations might well offer new insights into unsolved or poorly understood mysteries.

The definition and comparison of present-day patterns of genetic variation in South Asia not only offer unique insights for a deep study of human evolutionary history but also provide the necessary population–genetic background that constitutes an important prerequisite for understanding the genetics of complex traits.

**References**

1. Cameron D, Patnaik R, Sahni A. 2004. The phylogenetic significance of the Middle Pleistocene Narmada cranium from Central India. *Int J of Osteo* 14:419–447.
2. Sonakia A. 1984. The Skull-cap of Early Man and Associated mammalian fauna from narmada valley alluvium, Oasangabad area, Madhya Pradesh (India). *Records of the Geological survey of India* 113:159–172.
3. Singhvi AK, Deraniyagala SU, Sengupta D. 1986. Thermoluminescence dating of Quaternary red sand beds in Sri Lanka. *Earth Planat Sci Lett* 80: 139–144.
4. Kennedy KAR, Deraniyagala SU. 1989. Fossil remains of 28,000-year-old hominids from Sri Lanka. *Curr Anthropol* 30:394–399.
5. Dales GF. 1991. The phenomenon of the Indus civilization. In: Jansen M, Mulloy M, Urban G, editors. *Forgotten cities on the Indus: early civilization in Pakistan from the 8th to the 2nd millennia BC*. Mainz, Germany: Verlag Philipp von Zabern, pp 129–144.
6. Forster P. 2004. Ice ages and the mitochondrial DNA chronology of human dispersals: A review. *Philosophical Transactions of the Royal Society, London* 359:255–264.
7. James HVA, Petraglia MD. 2005. Modern Human Origins and the Evolution of Behaviour in the Later Pleistocene Record of South Asia. *Curr Anthropol* 46 suppl.
8. Westgate J, Shane R, Pearce N, Perkins W, Korissetar R, Chesner C, Williams M, Acharyya S. 1998. All Toba tephra occurrences across peninsular India belong to the 75,000 yr BP eruption. *Quaternary Research* 50:107–112.
9. Ambrose S. 1998. Late Pleistocene human population bottlenecks, volcanic winter, and differentiation in modern humans. *Journal of Human Evolution* 34:623–651.
10. Jorde LB, Bamshad M, Rogers AR. 1998. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* 20: 126–136.
11. Lewin R, Foley R. 2004. *Principles of Human Evolution*. Blackwell: Publishing Blackwell Science, 576 p.
12. Oppenheimer S. 2003. *Out of Eden: The peopling of the world*. London: Constable.
13. Richards M, Bandelt H-J, Kivisild T, Oppenheimer S. 2006. A model for the Dispersal of Modern Humans out of Africa. In: Bandelt HJ, Macauley V, Richards M, editors. *Human Mitochondrial DNA and the Evolution of Homo sapiens*. Berlin heidelberg: Springer-Verlag.
14. Klein RG. 2001. Southern Africa and modern human origins. *J of Anthropol Res* 57:1–16.
15. Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, et al. 1999a. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9:1331–1334.
16. Kivisild T, Kaldma K, Metspalu M, Parik J, Papiha S, et al. 1999b. The place of the Indian mtDNA variants in the global network of maternal lineages and the peopling of the Old World. In: Deka R, Papiha SS, editors. *Genomic Diversity*. New York: Kluwer Academic/Plenum Publishers, pp 135–152.
17. Kivisild T, Papiha SS, Rootsi S, Parik J, Kaldma K, et al. 2000. An Indian Ancestry: A Key for Understanding Human Diversity in Europe and Beyond. In: Renfrew C, Boyle K, editors. *Archaeogenetics: DNA and the Population History of Europe*. Cambridge: McDonald Institute for Archaeological Research, pp 267–275.
18. Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, et al. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet* 72:313–332.
19. Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM. 2001. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2:13.
20. Cann RL. 2001. Genetic clues to dispersal of human populations: Retracing the past from the present. *Science* 291:1742–1748.
21. Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, et al. 2003. Ethnic India: A genomic view, with special reference to peopling and structure. *Genome Res* 13:2277–2290.
22. Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, et al. 2006. A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc Natl Acad Sci USA* 103:843–848.



23. Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, et al. 2005a. Reconstructing the origin of Andaman Islanders. *Science* 13:996.
24. Thangaraj K, Chaubey G, Singh VK, Vanniarajan A, Ismail T, et al. 2006. In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. *BMC Genomics* 7:151.
25. Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. *The history and geography of human genes*. Princeton, NJ: Princeton University Press.
26. Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, et al. 2003. Human population genetic structure and inference of group membership. *Am J Hum Genet* 72:578–589.
27. Watkins WS, Prasad BV, Naidu JM, Rao BB, Bhanu BA, et al. 2005. Diversity and divergence among the tribal populations of India. *Ann Hum Genet* 69:680–692.
28. Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, et al. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics* 130:139–152.
29. Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, et al. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* 53:563–590.
30. Watson E, Forster P, Richards M, Bandelt H-J. 1997. Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61:691–704.
31. Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, et al. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034–1036.
32. Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, et al. 1999. Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat Genet* 23:437–441.
33. Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, et al. 1999. The Emerging Tree of West Eurasian mtDNAs: A Synthesis of Control-Region Sequences and RFLPs. *Am J Hum Genet* 64:232–249.
34. Endicott P, Macaulay V, Kivisild T, Stringer C, Cooper A. 2003. Reply to Cordaux and Stoneking. *Am J Hum Genet* 72:1590–1593.
35. Palanichamy M, Sun C, Agrawal S, Bandelt H-J, Kong Q-P, et al. 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.
36. Sun C, Kong Q-P, Palanichamy MG, Agrawal S, Bandelt H-J, et al. 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol Biol Evol* 3:683–690.
37. Excoffier L. 2002. Current Human demographic history: refining the recent African origin model. *Curr Opin in Genet & Dev* 12:675–682.
38. Tishkoff SA, Kidd KK. 2004. Implications of biogeography of human populations for 'race' and medicine. *Nat Genet* 36:S21–S27.
39. Mellars P. 2006. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* 313:796–800.
40. Torroni A, Richards M, Macaulay V, Forster P, Villems R, et al. 2000. mtDNA haplogroups and frequency patterns in Europe. *Am J Hum Genet* 66:1173–1177.
41. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67:1251–1276.
42. Yao YG, Kong QP, Bandelt H-J, Kivisild T, Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70:635–651.
43. Ingman M, Gyllensten U. 2003. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res* 13:1600–1606.
44. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, et al. 2003. Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 100:171–176.
45. Friedlaender J, Schurr T, Gentz F, Koki G, Friedlaender F, et al. 2005. Expanding Southwest Pacific mitochondrial haplogroups P and Q. *Mol Biol Evol* 22:506–517.
46. Merriwether D, Hodgson J, Friedlaender FR, Allaby R, Cerchio S, et al. 2005. Ancient mitochondrial M haplogroups identified in the Southwest Pacific. *Proc Natl Acad Sci USA* 102:13034–13039.
47. Passarino G, Semino O, Modiano G, Bernini LF, Santachiara Benerecetti AS. 1996a. 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.
48. Passarino G, Semino O, Bernini LF, Santachiara-Benerecetti AS. 1996b. Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. *Am J Hum Genet* 59:927–934.
49. 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.
50. Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, et al. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 98:10244–10249.
51. Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, et al. 2001. Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet* 68:537–542.
52. Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, et al. 2004. Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5:26.
53. Endicott P, Metspalu M, Kivisild T. 2006. Genetic evidence on modern human dispersals in South Asia: Y chromosome and mitochondrial DNA perspectives. In: Petraglia MD, Bridget A, editors. *The Evolution and History of Human Populations in South Asia*. SpringerLink publication (in press).
54. Boivin N. 2006. Anthropological, historical, archaeological and genetic perspectives on the origins of caste in South Asia. In: Petraglia MD, Bridget A, editors. *The Evolution and History of Human Populations in South Asia*. SpringerLink publication (in press).
55. Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, et al. 2001. Genetic evidence on the origins of Indian caste populations. *Genome Res* 11:994–1004.
56. Cordaux R, Aunger R, Bentley G, Nasidze I, Sirajuddin SM, et al. 2004a. Independent origins of Indian caste and tribal paternal lineages. *Curr Biol* 14:231–235.
57. Vivekananda S. 1985. *Swami Vivekananda on India and her Problems*. Calcutta: Advaita Ashram.
58. Majumder PP. 2001. Ethnic populations of India as seen from an evolutionary perspective; *J Biosci* 26:533–545.
59. Kumar V, Langsateh BT, Biswas S, Babu P, Rao TN, et al. 2006. Asian and Non-Asian Origins of Mon-Khmer- and Mundari-Speaking Austro-Asiatic Populations of India. *Am J Hum Biol* 18:461–469.
60. Cordaux R, Saha N, Bentley GR, Aunger R, Sirajuddin SM, et al. 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *European Journal of Human Genetics* 11:253–264.
61. Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, et al. 2004. Where West meets East: The complex mtDNA landscape of the Southwest and Central Asian corridor. *Am J Hum Genet* 74:827–845.
62. McElreavey K, Quintana-Murci L. 2002. Understanding inherited disease through human migrations: a south-west Asian perspective. *Community Genet* 5:153–156.
63. Cordaux R, Weiss G, Saha N, Stoneking M. 2004b. The northeast Indian passageway: a barrier or corridor for human migrations? *Mol Biol Evol* 21:1525–1533.
64. Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, et al. 2006. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central asian pastoralists. *Am J Hum Genet* 78:202–221.
65. Gutala R, Carvalho-Silva DR, Jin L, Yngvadottir B, Avadhanula V, et al. 2006. A shared Y-chromosomal heritage between Muslims and Hindus in India. *Hum Genet*, Nov; 120:543–551.
66. Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598–612.
67. Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, et al. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 114:127–148.
68. Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, et al. 2001. Origins and divergence of the Roma (gypsies). *Am J Hum Genet* 69:1314–1331.
69. Renfrew C. 1988. *Archaeology and language: the puzzle of Indo-European origins*. New York: Cambridge University Press .

70. Renfrew C. 1999. Time depth, convergence theory, and innovation in Proto-Indo-European. *J Indo-Eur Stud* 27:257–293.
71. Gurney OR. 1990. *The Hittites*. London/NY: Penguins.
72. Roux G. 1992. *Ancient Iraq*, 3rd ed. London: Penguins.
73. Dunstan WE. 1998. *The Ancient Near East*. NY/Orlando/London: Harcourt Brace College Publishers.
74. Kazanas ND. 2002. 'Indigenous Indoaryans and the RV'. *JIES* 30/3–4: 275–234.
75. Feuerstein G, Frawley D, Kak S. 1995. *In Search of the Cradle of Civilization*. Wheaton: Quest Books.
76. Sethna DK. 1992. *The Problem of Aryan Origins*, 2nd ed. Delhi: Aditya.
77. Anthony DW. 1995. Horse, wagon, and chariot: Indo-European languages and archaeology. *Antiquity* 69:554–565.
78. Gimbutas M. 1995. Primary and secondary homeland of the Indo-Europeans. *J Indo-European Studies* 13:185.
79. Renfrew C. 1987. *Archaeology and language: the puzzle of Indo-European origins*; *Archeol. And Laung*. London: Jonathan Cape.
80. Cavalli-Sforza LL. 1996. The spread of agriculture and nomadic pastoralism: insights from the genetics; linguistics and archaeology. In: Harris DR, editor. *The origins and spread of Agriculture and Pastoralism in Eurasia*. Washington, DC: Smithsonian Institution Press, pp 196–200.
81. Renfrew C. 1996. Languages families and the spread of farming. In: Harris DR, editor. *The origins and spread of agriculture and pastoralism in Eurasia*. Washington, DC: Smithsonian Institution Press, pp 70–92.
82. Blazek V. 1999. Elam: a bridge between Ancient Near East and Dravidian India? In: Roger Blench, Matthew Spriggs, editors. *Archaeology and Language IV. Language Change and Cultural Transformation*. London, New York: Routledge, pp 48–78.
83. Fuller D. 2003. *An Agricultural Perspective on Dravidian Historical Linguistics: Archaeological Crop Packages, Livestock and Dravidian Crop Vocabulary*. In: Bellwood P, Renfrew C, editors. *Examining the farming/language dispersal hypothesis*. Cambridge: The McDonald Institute for Archaeological Research.
84. Higham C. 1996. *The Bronze Age of Southeast Asia*. Cambridge: Cambridge Univ Press.
85. Kayser M, Brauer S, Weiss G, Schiefenhovel W, Underhill P, et al. 2003. Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet* 72: 281–302.
86. Thangaraj K, Sridhar V, Kivisild T, Reddy AG, Chaubey G, et al. 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. *Human Genet* 116:507–517.
87. Van driel G. 1999. A new theory on the origin of Chinese in Peter Bellwood and Ian Lilley, eds., *Indo-Pacific Prehistory Association Bulletin*, 18 (Melaka Papers, Volume 2) pp. 43–58.
88. Felsenstein J. 2004. *Inferring Phylogenies*. Sunderland, Massachusetts: Sinauer Associates.
89. Kivisild T, Shen P, Wall DP, Do B, Sung R, et al. 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172:373–387.
90. Misra VN, Nagar M. 1997. From tribe to caste: an ethnoarchaeological perspective; in *From tribe to caste* ed. Dev Nathan (Simla: Indian Institute of Advanced Study) pp 102–113.
91. Mishra VN. 2001. Prehistoric human colonization of India; *J Bio* 26 Supp.
92. Tambets K, Rootsi S, Kivisild T, Help H, Serk P, et al. 2004. The western and eastern roots of the Saami—the story of genetic “outliers” told by mtDNA and Y-chromosome. *Am J Hum Genet* 74:661–682.
93. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ. 2006. Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22:339–345.