Human Biology Review
www.humanbiologyjournal.com

Human Biology Review (ISSN 22774424) 8(4) 2019
Original scientific paper (Adak. pp. 339-366)
Modified and Accepted on September 10, 2019

DNA studies in India: a review

D.K. Adak

Citation: Adak DK. 2019. DNA studies in India: a review. Human Biology Review, 8 (4), 339-366.

Corresponding Author: Dipak Kumar Adak, Anthropological Survey of India, 27 Jawaharlal Nehru Road, Kolkata 700016, West Bengal, India. E-mail: adakdipak@gmail.com

ABSTRACT

Mitochondria of the cytoplasm possess some amount of DNA. One can find out enormous amount of DNA sequence variation within and between human populations. India with vast population diversity offer us ample opportunity to examine pattern of genetic structure and variation among the populations, migration and admixture in the past as well as geographic and ethnic affiliations of the groups. In this study an attempt has been made to review the DNA studies conducted in India.

mtDNA studies indicates deep-rooted maternal lineages for both caste and tribal populations of India and support single initial origin. The present population structure might have therefore resulted from subsequent admixture and drift. It is inferred that Y-Chromosome sharing between Northeast India and South China is due to Neolithic expansion via Northeast Indian Corridor. Autosomal marker studies shows genetic distances are not correlated with geographical distance.

Keywords: Population Variations. Studies on DNA. India

INTRODUCTION

Mendel's units of heredity were genes. The genes, responsible for various phenotypic expression and varied proteins, are present in the nucleus, and in the cytoplasmic materials like mitochondria and ribosomes.

The nucleus is very rich in basic genetic materials. Mitochondria of the cytoplasm also possess some amount of DNA. Though the ribosomes including the polysomes or polyribosomes do not have the basic genetic materials are however very rich in ribonucleic acid or RNA. The RNA are not primary genetic materials but the RNA helps in the phenotypic expressions of the genes by synthesizing proteins that are determined by their respective genes.

An enormous amount of research has been directed toward understanding DNA – what its structure is, how it duplicates itself in reproduction, and how it conveys or instructs the formation of a complete organism (Ember et al. 2003).

The advent of PCR techniques in the 1980s along with the use of restriction enzymes and later the use of sequencing method for identifying polymorphisms at the DNA level provided anthropologist with new and more powerful techniques and markers for testing different anthropological hypotheses (Tripathy et al. 2008). Techniques of DNA analysis have revealed that there exists enormous amount of DNA sequence variation within and between human populations. This variation may be of the type restriction fragment length

polymorphisms or certain base pair deletion or hypervariable DNA length polymorphisms or single nucleotide polymorphisms. These variations exists both in mitochondrial DNA and in the nuclear DNA (Barua 2007).

In the study of human variation, DNA sequences are being utilised in place of protein sequences of electrophoretic variation of proteins, immunogenetic traits like blood groups, and anthropometric traits. Hence, the study of DNA polymorphism (or DNA sequences) has become the most important part of biological anthropology. The handling and analysis of genomic DNA has become the routine work for the anthropological genetics (Barua 2007).

There have been attempt by anthropologists, linguistics, archaeologists and human geneticists to comprehend the vast diversity of the numerous tribes and castes; their past history; pattern of genetic variation among the populations; congruence between linguistic, geographic and ethnic affiliations of the groups. DNA polymorphisms reveal a population's genetic structure, migration and admixture in the past, susceptibility to illness and genetic causes of diseases. Human Genome, DNA sequencing offer an evidentiary alternative fossil based prehistoric reconstruction about origin and evolution of human beings. The uniparental inherited non-recombining haploid mtDNA and Y chromosome became an extremely powerful tool in phylogenetic studies (An.S.I. Kolkata: Unpublished).

DNA MARKERS: ADVANTAGES AND IMPORTANCE

Mitochondrial DNA

- Maternal inheritance
- Lack of recombination
- High substitution rate
- High copy number

Y-chromosomal DNA

- Paternally inherited
- Insertion/deletions
- Short Tandem Repeats (STRs)
- Single Nucleotide Polymorphism (SNPs)

Autosomal DNA

- Bi-parental inheritance
- Allelic distribution for population characterization
- Average heterozygosities
- Coefficient of gene differentiation
- STRs highly used in forensic

Source: Langstieh 2007

MITOCHONDRIAL DNA VARIATION IN INDIA

Studies on Hypervariable region I & II Sequences

One of the first studies based on mt DNA variation among the Indian population was attempted by Mountain et al (1995). He studied the three culturally divergent endogamous caste groups from coastal south western India namely Havik Brahmins, the Mukri and the Kadar. They estimated the time of divergence from common ancestors considering possible mutation rates. As per the molecular analysis, waves of migration into Indian sub-continent ranged from about 65 to 50 thousand years to 4 thousand years ago, pouring in various genetic materials and culturally adding all four language families, artefacts and cultivation of food crops (c.f. Ghoshmaulik 2000). Their study indicated that the Indian populations represented a major expansion possibly originated in Southern Asia.

Bamshad et al. (1996) analysed 36 Hindu men born in Andhra Pradesh who were unrelated matrilineally through at least 3 generations and who represent 4 caste populations: Brahmin (9), Yadava (10), Kapu (7) and Relli (10). They compared them with individuals from Africa (36), Asia (36) and Europe (36). A 200-base-pair segment of hypervariable segment 2 (HVS20 of the mtDNA control region was sequenced in all individuals. In the Indian castes, 25 distinct haplotypes are identified. Aside from the Cambridge reference sequence, only two haplotypes are shared between caste populations. Middle castes form a highly supported cluster in a neighbour-joining network. Mean nucleotide diversity within each caste is 0.015, 0.012, 0.011 and 0.012 for the Brahmin, Yadava, Kapu and Relli respectively. This study found that the effects of social structure on mtDNA variation are much greater than those on variation measured by traditional markers. The caste populations of Andhra Pradesh cluster more often with Africans than with Asians or Europeans, which is suggestive of admixture with African populations.

In a study of 250 individuals from 12 caste populations from northeastern Andhra Pradesh, Southern India, Bamshad et al. (1998) showed that differences in social rank between castes correspond to mtDNA distances, suggesting genetic stratification corresponding to social stratification and this they interpreted as due to the system of hypergamy, wherein women of lower caste hierarchy is allowed to marry men of higher castes (c.f. Tripathy et al. 2008).

Studies on RFLP based Bi-allelic Marker

Barnabas et al. (1996) studied mitochondrial DNA variation in 100 Indians, belonging to 14 different languages, using six restriction enzymes. They found that the Indian population is closer to Caucasians and has an admixture with Asians. The North Indian population appears to have a recent admixture of the Caucasian mtDNA types which is absent in the south. This supports the recent peopling of the Indo-European language-speaking people in India.

Passarino et al. (1996a) examined 70 individuals from Punjab for some mtDNA polymorphisms, namely, the RFLPs of the six classical enzymes (HpaI, BamHI, HaeII, MspI, AvaII, and Hin-cII) and for the sites AluI(7,025), DdeI(10,394), and AluI(10,397). They also examined the AluI(7,025) polymorphic site in 96 Indians from Uttar Pradesh and Andhra Pradesh and in 163 Mediterranean Caucasoid. Moreover, 30 Indian DdeI(10,394) Alu(10,397)

(++) mtDNAs were typed by the "high-resolution restriction analysis" with 14 endonucleases to estimate their divergence time. The results obtained from this study are the following:
(1) the RFLPs analysis has displayed some Caucasoid types as in Indians of Uttar Pradesh;
(2) the AluI(7,025) (-) allele, which defines the most frequent Caucasoid-specific lineage (haplogroup H) ranges from 18% to 45% in the Mediterranean Caucasoids, whereas it has shown low frequencies in Punjab (6.0%) and in Uttar Pradesh (1.8%) and was not found in Andhra Pradesh;
(3) the DdeI(lO,394)AluI(10,397) (+ +) haplotype, which although previously was considered an East Asian marker (haplogroup M) and was found very frequently in India, is also frequent in Punjab (27%); this frequency is, however, much lower than in Uttar Pradesh (49%) and in Andhra Pradesh (74%), and a gradient decreasing from south to north is therefore observed and
(4) the divergence time of the Indian DdeI(10,394)AluI(10,397) (++) mtDNAs has been estimated to be 30,250-60,500 years, a value that is compatible with that of the homologous East Asian lineage.

These results strongly support the hypothesis that the DdeI(10,394)AluI(10,397) (++) haplotype predated the Indo-European invasion and probably the split between proto-Indians and proto-Orientals. Its frequency cline well reflects the major influence of Indo-Europeans in the north and in the centre of India.

The concomitant presence of the two sites Ddel at 10,394 and Alul at 10,397 has been considered by Passarino et al. (1996b) in an East-Asian marker of ancient origin (it was also observed in Australians, Melanesians and Native Americans). Unexpectedly, they (Passarino et al. 1996b) found in more than 50% of Indians (133 Hindus and 30 Tribals) who had shown Caucasoid characteristics not only at nuclear DNA but also at mtDNA level. Thus, no longer it can be considered an exclusively East-Asian mtDNA feature.

The analysis of more than 200 Caucasoids, mainly from the Mediterranean basin, showed that it is only sporadically present in these people. Thus, it represents the first known marker, which distinguishes Indians from the other Caucasoids. They (Passarino et al. 1996b) are of the opinion that the lack of this marker in Indian mtDNA molecules carrying Caucasoid characteristics suggests that it predates the invasion of India by speakers of an Indo-European language and, if it is valid to extrapolate from Near Eastern data, the arrival in India of the farmers who spread the Dravidian language. If this polymorphism had a common origin in both Orientals and Indians, it should also predate the diversification between ancient Indians and Mongoloids.

Studies on 9 BP Deletion between COII and tRNA^{lys}

Mitochondrial DNA haplogroup B, is defined by a 9bp deletion in the intergenic region V of mtDNA, which is frequent in SE Asian populations, reaching almost fixation in some Polynesian populations. The 9bp motif is present between cytochrome oxidase II (CO II) and tRNA lysine (tRNA^{lys}) genes. Most individuals have two tandem copies of this 9bp motive. This 9bp deletion was earlier considered as a marker of Asian populations, subsequently found in most other population in varying frequencies (Tripathy et al. 2008).

Majumder (2001) studied some 30 populations from different regions of India. 8 mtDNA loci were screened. The 9 bp COII/$tRNA^{lys}$ intergenic length mutation revealed that all populations were monomorphic; no sampled individual showed the presence of the 9 bp deletion. The remaining seven RFLP loci were polymorphic.

Watkins et al. (1999) assayed for the presence of the intergenic COII/$tRNA^{lys}$ 9-bp deletion in human mtDNA in 646 individuals from 12 caste and 14 tribal populations of South India and compared them to individuals from Africa, Europe, and Asia. The 9-bp deletion is observed in four South Indian tribal populations namely the Irula, Yanadi, Siddi, and Maria Gond, and in the Nicobarese. Length polymorphisms of the 9-bp motif are present in the Santal, Khonda Dora, and Jalari, all of whom live in a circumscribed region on the eastern Indian coast. Phylogenetic analyses of mtDNA control region sequence from individuals with the 9-bp deletion indicate that it has arisen independently in some Indian tribal populations. Other 9-bp deletion haplotypes are likely to be of Asian and African origin, implying multiple origins of the 9-bp deletion in South India. These results demonstrate varying genetic affinities of different South Indian tribes to continental populations and underscore the complex histories of the tribal populations living in South Asia (Watkins et al. 1999).

Clark et al. (2000) analysed 9bp deletion in mtDNA of 898 individuals from 16 tribal populations of Northeast and Southern India. The frequency of 9bp deletion was found to be very low, just 0.8% in Northeast and 1.5% in Nilgiri Hills in South India. The 9bp deletion was reported only in 6 individuals and the sequences of these 6 formed 3 clusters from phylogenetic analysis, one cluster from Northeast showing similarity to Southeast Asian mtDNA types and the other two from South India were unique to India and showed no similarity with African mtDNA types (c.f. Tripathy et al. 2008).

To explore the origins and affinities of the Nicobar Islanders, Prasad et al (2001) analysed mtDNA hypervariable region 1 sequence data from 33 Nicobarese Islanders and compared their mtDNA haplotypes to those of neighbouring East Asians, mainland and island Southeast Asians, Indians, Australian aborigines, Pacific Islanders, and Africans. Unique Nicobarese mtDNA haplotypes, which include five Nicobarese mtDNA haplotypes linked to the COII/tRNALys 9-bp deletion, are most closely related to mtDNA haplotypes from mainland Southeast Asian Mon-Khmer-speaking populations (e. g., Cambodians). They are of the opinion that dispersal of southern Chinese into mainland Southeast Asia may have included a westward expansion and colonization of the islands of the Andaman Sea.

Thangaraj et al. (2005) found from the analysis of 3239 individuals of 58 endogamous populations of India that the frequency of 9bp deletion/insertion does not vary significantly among different and linguistic populations. They reported 20 independent origin of the 9bp deletion and insertion, which were not population specific. Frequency of 9bp deletion was found to be highest in the Austro-Asiatic speaking Nicobarese (45.8%). In contrast to Nicobarese, mainland Austro-Asiatic populations showed no incidence of 9bp deletion, which points towards an independent origin of Nicobarese and the mainland Austro-Asiatic populations.

mtDNA 9-bp deletion/insertion polymorphism was analysed among 1686 samples from 31 tribal populations of India by Kumar et al. (2006). A comparative analysis of their results with the existing data suggests multiple origins of Austro-Asiatic tribes in India, and particularly the Asian and non-Asian origins of the Mon-Khmer and Mundari populations. This study also identified a novel subclade of haplogroup B in the Mon-Khmer Khasi tribes that distinguishes them from the Nicobarese, indicating two different waves of migration of the Mon-Khmer tribes in India.

Studies on Haplogroup:

A commonly held hypothesis, suggests a massive Indo-Aryan invasion to India some 4,000 years ago and recent limited analysis of maternally inherited mitochondrial DNA (mtDNA) of Indian populations has been interpreted as supporting this concept. This interpretation is questioned by Kivisild et al. (1999). They found an extensive deep late Pleistocene genetic link between contemporary Europeans and Indians, provided by the mtDNA haplogroup U, which encompasses roughly a fifth of mtDNA lineages of both populations. Their estimate for this split is close to the suggested time for the peopling of Asia and the first expansion of anatomically modern humans in Eurasia and likely pre-dates their spread to Europe. Only a small fraction of the 'Caucasoid-specific' mtDNA lineages found in Indian populations, which can be ascribed to a relatively recent admixture (Kivisild et al. 1999).

Mitochondrial DNA (mt DNA) variation of 23 ethnic populations of India drawn from diverse cultural, linguistic and geographical backgrounds are presented by Roychoudhury et al. (2000). There is extensive sharing of a small number of mtDNA haplotypes, reconstructed on the basis of restriction fragment length polymorphisms, among the populations. This shows that Indian populations were founded by a small number of females, possibly arriving on one of the early waves of out-of-Africa migration of modern humans. The authors found that the Asian-specific haplogroup M is in high frequency in most populations, especially tribal populations and Dravidian populations of Southern India. They also found that populations in which the frequencies of haplogroup M are relatively lower show higher frequencies of haplogroup U. These populations are primarily caste populations of northern India. This finding is indicative of a higher Caucasoid admixture in northern Indian populations. By examining the sharing of haplotypes between Indian and south-east Asian populations the authors have provided evidence that south-east Asia was peopled by two waves of migration, one originating in India and the other originating in southern China.

Using a set of autosomal DNA markers, mtDNA restriction-site polymorphisms (RSPs) and mtDNA hypervariable segment-1 (HVS-1) sequence polymorphisms Roychudhury et al. (2001) conducted a genetic study among the tribal groups of India belonging to three major language groups viz. Austro-Asiatic (Santal=21, Munda=7 and Lodha=32), Dravidian (Muria=49, Kota=45, Kurumba=54, Irula=50) and Tibeto-Burman (Tipperah=51). They reconstructed mtDNA RSP haplotypes and found that there is extensive haplotype sharing among all tribal populations. However, there is very little sharing of mtDNA HVS-1 sequences across populations, and none across language groups. Haplogroup M is ubiquitous, and the subcluster U2i of haplogroup U occurs in a high frequency. Analyses of Roychoudhury et al. (2001) on haplogroup and HVS-1 sequence data provides evidence in support of the hypothesis that the Austro-Asiatic speakers are the most ancient inhabitants of India. Their data also support the earlier finding that some of the western Eurasian haplogroups found in India may have been present in India prior to the entry of Aryan

speakers. However, they do not find compelling evidence to support the theory that haplogroup M was brought into India on an "out of Africa" wave of migration through a southern exit route from Ethiopia. On the contrary, data of Roychoudhury et al. (2001) raise the possibility that this haplogroup arose in India and was later carried to East Africa from India.

Kivisild et al (2003) examined two tribal groups from southern India- the Chenchus and Koyas for variation in mtDNA, Y chromosome and one autosomal loci. The results were compared with six caste groups from different parts of India, as well as with western and central Asians. The study suggests that Indian caste and tribal populations are derived largely from the same genetic heritage of Pleistocene southern and western Asians and have received limited gene flow from external regions since the Holocene. The findings suggest that these southern Asian Pleistocene coastal settlers from Africa would have inoculated for the subsequent differentiation of the distinctive eastern and western Eurasian gene pools.

To explore the impact of West Eurasians on contemporary Indian caste populations Bamshad et al (2001) compared mtDNA (400 bp of hypervariable region 1 and 14 restriction site polymorphisms) and Y-chromosome (20 biallelic polymorphisms and 5 short tandem repeats) variation in approximately 265 males from eight castes of different rank to approximately 750 Africans, Asians, Europeans, and other Indians. They found that 20%-30% of Indian mtDNA haplotypes belong to West Eurasian haplogroups, and the frequency of these haplotypes is proportional to caste rank. The highest frequency of West Eurasian haplotypes being found in the upper castes. In contrast, for paternally inherited Y-chromosome variation each caste is more similar to Europeans than to Asians. Moreover, the affinity to Europeans is proportionate to caste rank, the upper castes being most similar to Europeans, particularly East Europeans. According to them these findings are consistent with greater West Eurasian male admixture with castes of higher rank. Nevertheless, the mitochondrial genome and the Y chromosome each represents only a single haploid locus and is more susceptible to large stochastic variation, bottlenecks, and selective sweeps. To increase the power of analysis, Bamshad et al. (2001) assayed 40 independent, biparentally inherited autosomal loci (1 LINE-1 and 39 Alu elements) in all of the caste and continental populations (approximately 600 individuals). Analysis of these data demonstrated that the upper castes have a higher affinity to Europeans than to Asians, and the upper castes are significantly more similar to Europeans than are the lower castes. Collectively, all five datasets show a trend toward upper castes being more similar to Europeans, whereas lower castes are more similar to Asians. They conclude that Indian castes are most likely to be of proto-Asian origin with West Eurasian admixture resulting in rank-related and sex-specific differences in the genetic affinities of castes to Asians and Europeans.

370 bp of the first hypervariable region of the mtDNA control region is analysed by Cordaux and Stoneking (2003) in 752 individuals from 17 tribal and four non-tribal groups from the Indian subcontinent. South Indian tribes showed reduced diversity and large genetic distances. Whereas, northern groups exhibited more diversity. It is revealed from phylogenetic analyses that southern and northern groups have related mtDNA sequences. The Indian mtDNA gene pool appears to be more closely related to the east Eurasian gene pool than the west Eurasian one. They suggested that caste and tribal groups are genetically similar with respect to mtDNA variation.

Baig et al. (2004) exmined mtDNA variation among seven communities belong to tribes and castes from western Maharashtra. They found that nucleotide diversity, gene diversity and average mismatches were of the same magnitude. It is seen from mtDNA haplogroups that both caste and tribal populations share same branches of the trees and the maternal lineages have

their roots in early late Pleistocene. Side by side, Thanseem et al. (2006) did not find significant difference between Indian caste and tribal populations for mtDNA; still higher frequency of west Eurasian specific haplogroups were found in the higher castes especially from north western part of India.

The south-western and central Asian corridor has played a pivotal role in the history of humankind, witnessing numerous waves of migration of different peoples at different times. To evaluate the effects of these population movements on the current genetic landscape 910 mtDNAs were analysed by Quintana-Murci et al. (2004). These represented 23 populations of Iranian plateau, Indus Valley and Central Asia. Indus Valley population comprised of populations from Pakistan and a group of Gujratis from India. The study showed that populations located west of the Indus Valley mainly harbour mtDNA of western Eurasian origin, whereas those inhabiting the Indo-Gangetic region and Central Asia present substantial proportions of lineages that can be allocated to three different genetic components of western Eurasian, eastern Eurasian and south Asian origin.

Recent advances in the understanding of the maternal and paternal heritage of south and southwest Asian populations have highlighted their role in the colonization of Eurasia by anatomically modern humans. Further understanding requires a deeper insight into the topology of the branches of the Indian mtDNA phylogenetic tree, which should be contextualized within the phylogeography of the neighboring regional mtDNA variation. Accordingly, Metspalu et al. (2004) analyzed mtDNA control and coding region variation in 796 Indian (including both tribal and caste populations from different parts of India) and 436 Iranian mtDNAs. This study defined four new Indian-specific haplogroup M sub-clades. These, in combination with two previously described haplogroups, encompass approximately one third of the haplogroup M mtDNAs in India. The analysis of the Iranian mtDNA pool revealed patterns of limited reciprocal gene flow between Iran and the Indian sub-continent and allowed the identification of different assemblies of shared mtDNA sub-clades. Since the initial peopling of South and West Asia by anatomically modern humans, when this region may well have provided the initial settlers who colonized much of the rest of Eurasia, the gene flow in and out of India of the maternally transmitted mtDNA has been surprisingly limited. Analysis of the mtDNA haplogroups, which are shared between Indian and Iranian populations and exhibit coalescence ages corresponding to around the early Upper Paleolithic, indicates that they are present in India largely as Indian-specific sub-lineages. In contrast, other ancient Indian-specific variants of M and rare very rare outside the sub-continent.

Sahoo and Kashyap (2006) examined variation in mtDNA hypervariable sequence (HVS) I and II, eight Y-chromosome short tandem repeats (Y-STRs), and lineage-defining mutations diagnostic for Indian and Eurasian specific haplogroups among seven caste and tribal populations of Orissa. They found, from the data of mtDNA, the hierarchical association with the Indo- European speakers of Eastern Europe.

Many efforts based on complete mitochondrial DNA (mtDNA) genomes have been made to depict the global mtDNA landscape, but the phylogeny of Indian macrohaplogroup M has not yet been resolved in detail. To fill this lacuna Sun et al. (2006) took the same strategy as in their analysis of Indian mtDNA macrohaplogroup N. They selected 56 mtDNAs from over 1,200 samples across India for complete sequencing, with the intention to cover all Indian autochthonous M lineages. As a result, the phylogenetic status of previously identified haplogroups based on control-region and/or partial coding-region information, such as M2, M3, M4, M5, M6, M30, and M33, was solidified or redefined. Moreover, seven novel basal M haplogroups (viz., M34-M40) were identified, and yet another five singular branches of the M

phylogeny were discovered by them. The comparison of matrilineal components among India, East Asia, Southeast Asia, and Oceania at the deepest level yielded a star-like and non-overlapping pattern, reflecting a rapid mode of modern human dispersal along the Asian coast after the initial "out-of-Africa" event.

To study the phylogenetic relationship of Indian and Western Eurasian mtDNA Palanichamy et al. (2004) did complete sequencing of 75 individuals belonging to macrohaplogroup N. The study showed that Indian N macrohaplogroup consists of some lineages, which are as deep rooted as the western Eurasian mtDNA lineages. In addition, some typical European haplogroups like H, V, K, U5, J and T provide an entry time into India, which is less than 11.5 kya. It was concluded that these results along with the evidence for M macrohaplogroup suggest three founder mtDNA lineages migrating into the subcontinent at different times.

Studies, which focus on northeastern populations largely, conclude that they differ from mainland populations and show affinity to south East Asian populations. In a study of biallelic and five short tandem-repeat Y chromosome markers and mtDNA hypervariable region I sequence variation in 192 northeast Indians Cordaux et al. (2004a) found that northeast Indian mtDNAs consistently show strikingly high homogeneity among groups and strong affinities to East Asian groups. They detected virtually no mtDNA admixture between Northeast and other Indian groups. Northeast Indian groups are also characterized by a greatly reduced Y chromosome diversity, which contrasts with extensive mtDNA diversity. Based on both the mtDNA and Y chromosome results they conclude that there is a strong evidence for a genetic discontinuity between Northeast Indian groups and other Indian groups, hence the Northeast Indian passageway acted as a geographic barrier rather than as a corridor for human migrations between the Indian subcontinent and East/Southeast Asia. Given the representation of populations in this study, the above conclusions are to be taken with a pinch of salt. In this context, it is important to note that the Northeast Indian studies hither to neglect certain important tribal populations like Khasis who speak Khasi-Khmuic, an Austro-Asiatic language in the midst of predominantly Indo-European speakers. In a recent study, Reddy et al. (2007) found mtDNA evidence on relic genetic link between Indian and Northeast Indian populations, contrary to Cordaux et al (2004a) inference, besides a strong Y chromosomal connection between the Mundari tribes from Central India, Khasi and Monkhmer Nicobarese (the three linguistic subfamilies) as well as with their linguistic counterparts from the Southeast Asian region (Kumar et al. 2007) (c.f. Tripathy et al. 2008).

There have been a few studies focusing on the affinities of Indian Muslim Populations, the largest religious minority of India. Terros et al. (2007) studies the mtDNA variation in the two Muslim sects from the Northern Indian province of Uttar Pradesh, the Sunni and Shia. A comparison of this data to that from Middle Eastern, Central Asian, North East African, and other Indian groups revealed that, at the mtDNA haplogroup level, both of these Indian-Sunni and Shia populations are more similar to each other and to the other Indian groups than to those from the other regions. These two Muslim sects exhibit a conspicuous absence of West Asian mtDNA haplogroups suggesting that their maternal lineages are of Indian origin (c.f. Tripathy et al. 2008).

Studies on Y-Chromosome variation:

The paternally inherited Y-Chromosome provides insights into the origin, history and migratory pathways of the male lineages and helps reconstruction the evolutionary history of the populations. In fact one of the first papers describing Y-Chromosome variation compared

relative Y-Chromosome and mtDNA distances among the hierarchically stratified castes in Andhra Pradesh (Bamshad et al. 1998). In contrast to mtDNA, Y-Chromosome distances showed no correlation with social rank. This was interpreted as due to lack of male gene flow between castes of different social rank. Bamshad et al. (1998) concluded that the genetic stratification of the Hindu caste system is driven by the social mobility of women in a patrilocal society and social laws of hypergeny played a role in it. Bhattacharya et al. (1999) found that different ethnic groups of India represented in this study primarily by certain Indo-European and Austro-Asiatic linguistic groups harbour disjoint sets of haplotypes. In addition, there was a significant haplotype variation between castes and tribes. The authors interpreted this as due to lack of male gene flow across ethnic groups of India. Nevertheless, it may be pertinent to note that the populations included in this study were geographically, linguistically and ethnically so disperse that they had no possibility for exchange of genes either historically or currently. Based on Y-Chromosome data, Sahoo and Kashyap (2006) reported that Y-Chromosome data suggest genetic distances of the populations are not correlated with their position in the caste hierarchy, which was in contrast to their mtDNA data. Though the genetic distances based on Y were not corrected with social rank. The higher caste groups were closer to the Indo-European speakers of Eastern Europe. Basu et al. (2003) observed significant differences in the frequency of Y-Chromosome haplogroups, in contrast to those of mtDNA, between the tribal and caste groups. Also the frequencies of Central Asian haplogroups are higher in the caste than that the tribal populations. In a study of Y-Chromosome variation among 4 caste populations, 3 tribal populations and Siddis from Andhra Pradesh. Ramana et al. (2001) found that Y-Chromosome haplotypes are unique to castes and tribes, so that they could be distinguished on the basis of haplotypes. But there were certain haplotypes which were present in all the caste and tribal groups. Ramana et al. (2001) concluded this as evidence of male gene flow across castes and tribes in the region. This conclusion is in contrast to what has been proposed by Bamshad et al. (1998) and Bhattacharya et al. (1999) (c.f. Tripathy 2008).

Linguistic evidence suggests that West Asia and Central Asia have been the two major geographical sources of genes in the contemporary Indian gene pool. To test the nature and extent of similarities in the gene pools of these regions Mukherjee et al. (2001) collected DNA samples from four ethnic populations of northern India, and have screened these samples for a set of 18 Y-chromosome polymorphic markers (12 unique event polymorphisms and six short tandem repeats). These data from Indian populations have been analysed in conjunction with published data from several West Asian and Central Asian populations. Analysis of Mukherjee et al. (2001) revealed traces of population movement from Central Asia and West Asia into India. Two haplogroups, HG-3 and HG-9, which are known to have arisen in the Central Asian region, are found in reasonably high frequencies (41.7% and 14.3% respectively) in the study populations. The ages estimated for these two haplogroups are less in the Indian populations than those estimated from data on Middle Eastern populations. A neighbour-joining tree based on Y-haplogroup frequencies shows that the North Indians are genetically placed between the West Asian and Central Asian populations. This is consistent with gene flow from West Asia and Central Asia into India.

Unlike mtDNA variation, Y-chromosome variation in each of eight castes from the state of Andhra Pradesh was found to be more similar to East Europeans than to Asians (Bamshad et al. 2001). The affinity of different Indian castes to Europeans was observed to be proportionate to caste rank. Based on mtDNA variation. They found that Indian castes are

likely to be of proto-Asian in origin and Y-chromosome variation indicates west Eurasian admixture, which is proportionate to the caste rank.

Similar findings of both mtDNA and Y-chromosome variation were reported from some studies. Findings of Kivisild et al. (2003) report that the major Indian Y-chromosome haplogroups H, L and R2 occur in both the castes and tribal populations, which are rarely found outside the subcontinent. Haplogroup R1a, which was earlier associated with Indo-Aryan invasion, was found in high frequency in Punjab as well as in the Chenchu tribe in south India. Therefore, the results of both mtDNA and Y-chromosome variation suggest a common genetic heritage of Indian castes and tribes.

Three tribal populations from Andhra Pradesh was analysed by Thanseem et al. (2006). The results of the study were compared with other populations. It was found that in contrast to mtDNA, Y-chromosome variation in India is distinct among caste and tribal populations. The lower castes are characterized with closer affinity to tribal populations than to upper castes and the frequencies of deep rooted Y-haplogroups were higher in the lower castes and tribes in comparison with upper castes. This suggest the tribal origin of the lower castes.

Thirty-eight single-nucleotide polymorphic markers was typed by Sahoo et al. (2006) in 936 Y-chromosomes representing 32 tribal and 45 caste groups from all four major linguistic groups of India. This study suggests that the recent external contribution to Dravidian and Hindi speaking caste groups has been low and rule out recent major influx from North and West of India.

Sengupta et al. (2006) examined 69 Y-chromosome SNPs and 10 microsatellite markers from a large set of geographically, socially, linguistically and ethnically diverse groups of South Asia. The study found that the influence of Central Asia on the pre-existing gene pool was minor. The ages estimated for the accumulated microsatellite variation in the majority of Indian haplogroups exceeded 10,000-15,000 years thus ruling out the pronounced recent genetic input from Central Asia. The findings also support the deep antiquity of Indian populations.

Austro-Asiatic tribal groups possess high frequencies of Y-chromosomal haplogroup K. This is found in higher frequencies in Chinese and Southwest Asian populations (Basu et al. 2003). According to them Austro-Asiatic tribal populations entered India first from the Northwest corridor and much later some of them through Northeastern corridor. Kumar et al. (2007) analysed a battery of relevant Y-chromosome SNPs and 20 STRs among 25 Indian Austro-Asiatic tribes, including the transitional ones, and compared with 214 relevant populations from Asia and Oceania to trace the origin and historic expansion of Austro-Asiatic groups of India. Strong paternal genetic link was found not only among the sub-linguistic groups of the Indian Austro-Asiatic populations but also with those of South East Asia. Maternal link based on mtDNA was not that apparent, however. Results also indicate that the haplogroup O-M95 had originated in the ancestors of Indian Austro-Asiatic populations ~65,000 yrs BP and was further carried to Southeast Asia via the Northeast Indian corridor. The conclusion of Kumar et al. (2007) and Basu et al. (2003) are different from that of Cordaux et al. (2004a) who had proposed that Northeast India acted as barrier. Nevertheless, Kumar et al. (2007) differ in the direction of migration for Austro-Asiatic populations through the Northeast corridor (c.f. Tripathy et al. 2008).

Are the Hindu castes largely derived from Indian local populations (i.e. tribal groups) or from recent immigrants to India? Archaeological and linguistic evidence support the latter hypothesis, whereas recent genetic data seem to favour the former hypothesis. Keeping this in view Cordaux et al. (2004b) analyse the most extensive dataset of Indian caste and tribal Y

chromosomes to date. They found that caste and tribal groups differ significantly in their haplogroup frequency distributions; caste groups are homogeneous for Y chromosome variation and more closely related to each other and to central Asian groups than to Indian tribal or any other Eurasian groups. The study conclude that paternal lineages of Indian caste groups are primarily descended from Indo-European speakers who migrated from central Asia approximately 3,500 years ago. Conversely, paternal lineages of tribal groups are predominantly derived from the original Indian gene pool. Cordaux et al. (2004b) also provide evidence for bidirectional male gene flow between caste and tribal groups. In comparison, caste and tribal groups are homogeneous with respect to mitochondrial DNA variation, which may reflect the sociocultural characteristics of the Indian caste society.

Because of the widespread phenomenon of patrilocality, it is hypothesized that Y-chromosome variants tend to be more localized geographically than those of mitochondrial DNA (mtDNA). Empirical evidence confirmatory to this hypothesis was subsequently provided among certain patrilocal and matrilocal groups of Thailand, which conforms to the isolation by distance mode of gene diffusion (Kumar et al. 2006b). They have tested the universality of this hypothesis by analyzing Y-chromosome and mtDNA data in three different sets of Indian populations that follow endogamy to varying degrees. The study showed that the Indian patrilocal and the matrilocal groups is not confirmatory to the sex-specific variation observed among the tribes of Thailand. The results indicate spatial instability of the impact of different cultural processes on the genetic variability, resulting in the lack of universality of the hypothesized pattern of greater Y-chromosome variation when compared to that of mtDNA among the patrilocal populations.

Redd et al. (2002) showed addition DNA evidence in support of Huxley's hypothesis of an Indian-Australian connection using SNPs and STRs on the non-recombining portion of the Y-chromosome (NRY). Phylogenetic analysis of STR variation associated with a major Australian SNP lineage indicated tight clustering with southern Indian/Sri Lankan Y chromosomes. Estimates of the divergence of these Indian and Australian chromosomes overlap with important changes in the archaeological and linguistic records in Australia. These results provide strong evidence for an influx of Y chromosomes from the Indian subcontinent to Australia that may have occurred during the Holocene (c.f. Tripathy 2008).

Arab forces conquered the Indus Delta region in 711 A.D. and, although a Muslim state was established there, their influence was barely felt in the rest of South Asia at that time. By the end of the tenth century, Central Asian Muslims moved into India from the Northwest and expanded throughout the subcontinent. Muslim communities are now the largest minority religion in India, comprising more than 138 million people in a predominantly Hindu population of over one billion. It is unclear whether the Muslim expansion in India was a purely cultural phenomenon or had a genetic impact on the local population (Gutala et al. 2006). To address this question from a male perspective, Gutala et al (2006) typed eight microsatellite loci and 16 binary markers from the Y chromosome in 246 Muslims from Andhra Pradesh, and compared them to published data on 4,204 males from China, Central Asia, other parts of India, Sri Lanka, Pakistan, Iran, the Middle East, Turkey, Egypt and Morocco. They find that the Muslim populations in general are genetically closer to their non-Muslim geographical neighbours than to other Muslims in India, and that there is a highly significant correlation between genetics and geography (but not religion). Findings of Gutala et al. (2006) indicate that, despite the documented practice of marriage between Muslim men and Hindu women, Islamization in India did not involve large-scale replacement of Hindu Y chromosomes. The Muslim expansion in India was predominantly a cultural change and was not accompanied by significant gene flow, as seen in other places, such as China and Central Asia.

Studies on Autosomal variation:

As against mtDNA and Y-chromosome, which show a simple pattern of inheritance, the autosomal DNA has not been a favourite in the studies of peopling of India, especially those which focused on the migration of humans. Still quite a few studies have used autosomal markers in understanding the genetic affinities of Indian populations. A lot of data have also been generated on many autosomal (and Y as well) STR/microsatellite markers by Central Forensic Science Laboratory, Kolkata and others. Many published articles from this institution and others have focused on forensic utility of specific markers in studied populations (c.f. Tripathy 2008).

DNA samples from 396 individuals belonging to 14 ethnic groups of India were analysed by Majumder et al. (1999) for 8 human specific polymorphic insertion/deletion loci. They observed that geographically closer populations showed greater affinity than populations with socio-cultural similarity. But Roychoudhury et al. (2001) found genomic affinity to show good association with linguistic affinity for autosomal loci. Side by side, Viswanathan et al. (2004) found that genetic distances based on 24 autosomal markers were not correlated with geographic distances for 5 Dravidian speaking tribal populations from Nilgiri Hills, Tamil Nadu. This study also found that the Indian populations were closely related to each other, regardless of the fact that some tribal populations showed Negrito phenotypic characteristics. It was observed by Basu et al (2003) that the clustering, on the basis of mtDNA, Y-chromosome and autosomal markers, was not strictly consistent with their social, geographical or linguistic affiliations, except the Tibeto-Burman speakers of Northeast India, which formed a separate cluster. It is evident from the analysis of population structure that Indo-European speakers and Dravidian speakers are the most similar for autosomal data, though they stand geographically wide apart.

Autosomal STRs and population structure at micro and macro levels:

Ashma and Kashyap (2003) demonstrated social hierarchical relationship among four caste groups from Bihar, for 15 microsatellite markers. Kashyap et al. (2004a) reported that the clustering pattern corresponds with the spatial and ethnic affiliation of the eight population groups from West Bengal and Manipur for 22 autosomal loci. Based on twelve microsatellite markers. Krithika et al. (2006) found that genetic affinity was correlated with geographical distance for 14 Tibeto-Burman populations. Bamshad et al. (2001) using 40 autosomal markers found highly significant hierarchical stratification of the caste populations of Andhra Pradesh. However, Reddy et al. (2005) typed 9 AmpF/STR Profiler Plus loci on 948 individuals from 27 populations (both caste and tribal) from Southern Andhra Pradesh and the inferences were somewhat contradictory. They found allele frequency distributions are fairly uniform across the populations of southern Andhra Pradesh. The caste groups showed the largest genetic distances with tribes when compared to the mutual distances among them, suggesting genetic isolation of the tribes and castes. There is also a meek trend of increase in genetic distances with the increasing hierarchy of populations and this they inferred as probably due to unauthorized male gene flow rather than hypergamy of females. Further they reported lack of any pattern in the population clustering based on ethnohistoric or geographical affiliations. But when compared with other Indian and continental populations, the studied Andhra populations form a single and compact cluster. Thus, at an all India level, irrespective of the social hierarchy these populations from a single geographical area are genetically homogenous to each other. Langstieh et al. (2004) analysed 9 AmpFL STR loci for 932 Chromosomes from 9 populations out of which 7 were subpopulations of Austro-Asiatic Mon-Khmer speaking Khasi, one neighbouring Tibeto-Burman speaking Garo and an intermediate population, Lyngngam. The

different analyses revealed lack of clear differentiation and clustering pattern in these populations. The authors interpret the reduced microsatellite diversity as partly due to matriliny/matrilocality practised by these populations. When these populations are compared with other Indian, Southeast Asian, and other continental populations, the Meghalaya populations form a compact cluster separated from other populations, suggesting genetic identity of these populations. Overall, the observation of Reddy et al. (2005) showed that geographically closer populations are genetically closer is true for different linguistic populations of Meghalaya as well. However, these populations show greater proximity to the other Mongoloid populations from the Southeast/East Asian countries. Similar findings were apparent in case of Northeast Indian populations in general vis-à-vis the other Indian and East/Southeast Asian populations (Dutta et al. 2002). In another study from Andhra Pradesh, Reddy et al. (2001a) used 13 STR loci in seven populations of a substructured Golla caste from Chittor district of Andhra Pradesh. Genetic distance measures revealed clusters of populations that are consistent with the known ethnohistorical and geographical backgrounds of the groups. Similar results were obtained by Reddy et al. (2001b) based on only 3 of the above 13 STR loci among the Gollas. Thus, the patterns observed at a local level, within a sub-structured caste population, were not apparent when many tribal and caste populations from heterogenous geographic background were studied (c.f. Tripathy et al. 2008).

In a study, Kivisild et al (2003) found that frequency of autosomal haplogroups namely MX1 and Ch21 distinguished two tribes, the Koyas and Chenchus from South India, along with other Indian castes, from the European and Eastern Asian populations. It is interesting that the result corroborates with their mtDNA and Y-Chromosome based conclusion, which suggest common ancient genetic heritage for Indian caste and tribal populations.

Rajkumar and Kashyap (2004) described polymorphism at 15 autosomal STR in four endogamous populations from Karnataka on the Southwest coast of India and their results indicated common ancestry for the four diverse populations of Karnataka. In a similar study (Sahoo and Kashyap 2005) among seven populations belonging to two major ethnic groups and different linguistic families inhabiting the same geographical area suggested genetic heterogeneity although still the contemporary caste and tribal groups formed distinct clusters in both Principal Component plot and neighbour joining tree. Congruent to this, Watkins et al. (2005) study of 45 unlinked autosomal STR loci portray relative closeness of Indian tribal and caste populations between them than no other major ethnic groups from outside. The shared phenotypic characteristics of some tribes with African were not reflected in their genetic composition. South Indian populations showed lower within heterozygosities compared to the Northeast Indian populations. Gaikwad et al. (2006) reported Microsatellite diversity among four Proto-Australoid tribes from west-central India. The relationship of these tribes with neighbouring tribal and caste populations were studied. Overall, the results suggested that the Proto-Australoid tribal populations were genetically differentiated from castes of similar morphology, suggesting different evolutionary mechanisms and their intensities upon these populations. On the other hand, Kashyap et al. (2004b, 2006) assessed the variation in microsatellite markers among 3522 individuals belonging to 54 endogamous Indian populations representing all major ethnic, linguistic and geographic groups and observed that the distribution of the most frequent alleles were uniform across populations, revealing an underlying genetic similarity, as observed by Reddy et al. (2005) in an array of hierarchical caste and tribal groups of southern Andhra Pradesh. Autosomal microsatellite markers detected no evidence of general clustering of population groups based on ethnic, linguistic, geographic or socio-cultural affiliations. These conclusions are in agreement with conclusion of common genetic heritage of Indian populations based on mtDNA and Y-Chromosome (Kivisild et al. 2003; Roychoudhury et al. 2000; Basu et al. 2003) (c.f. Tripathy et al. 2008).

Studies on Andaman and Nicobar populations:

The issue of the origin of Andaman and Nicobar tribes has always been interesting to Anthropologists. These tribes have been proposed to provide footprints of migration of early wave of modern humans to Australia (Redd et al. 2002; Macaulay 2005). Few studies have focused on the molecular genetic features of populations from Andaman and Nicobar islands. Analyzing mtDNA sequences and RFLP polymorphisms, Y-Chromosome biallelic marker and microsatellites in the populations of Andaman and Nicobar Islands, Thangaraj et al. (2003) observed low genetic variation among them. The closer genetic affinity of Andamanese to the Asians than to African populations suggests that they are the descendants of the early Paleolithic colonizers of South-East Asia. Thangaraj et al. (2005) reported low haplotype diversity among the 9bp samples in the Nicobarese, which suggests recent founder effect and origin in China via Cambodia and Thailand. Complete mtDNA sequences of five Onge, five Great Andamanese, and five Nicobarese individuals analysed by Thangaraj et al. (2005) suggest that the two ancient maternal lineages, M31 and M32 in the Onge and the great Andamanese have evolved in the Andaman Islands independently and not derived from the South and/or Southeast Asian populations. These lineages have been suggested as probably isolated since the initial colonization of the northern coastal areas of the Indian Ocean by anatomically modern humans, in the trail of their out-of-Africa migration ~50-70kya. In contrast, the Nicobarese show a close genetic relation to the populations of Southeast Asia, suggesting their recent arrival from the east during the past 18 thousand years. Concurrent to this, Thangaraj et al. (2006) based on autosomal STR loci, concluded that the Andaman "Negrito" populations do not show particular affinities either to the African or Indian populations whereas the Nicobarese show close affinities with the Southeast Asian populations, suggesting their recent colonization of the Islands as reflected by the mtDNA data. Prasad et al. (2001) analysed mtDNA hypervariable region 1 sequence data from 33 Nicobarese Islanders and compared their mtDNA haplotypes to those of neighbouring East Asians, Indians, Australian aborigines, Pacific Islanders, and Africans. They found that the unique Nicobarese mtDNA haplotypes were most closely related to Southeast Asian Mon-Khmer speaking Cambodian populations. They concluded that the Nicobarese population is the result of westward expansion of the Southeast Asian populations. Trivedi et al. (2006) analysed the mitochondrial, Y-Chromosomal and autosomal gene pools of contemporary Shompen, which belongs to Mon Khmer linguistic subfamily of the Austro-Asiatics. Overall, as was the case with other tribes of these islands, this tribe shows low genetic diversity. Mitochondrial DNA sequence analyses revealed the presence of two haplo-groups of R lineage: B5a, and a newly defined clade, R12, Y-chromosomal analysis suggest predominant occurrence of a single lineage, O-M95, which was found most predominantly in Mundari Austro-Asiatic tribes and Nicobarese as well as among the other Austro-Asiatics found across Asia, with the different types of genetic markers analysed, the Shompen exhibits varying levels of genetic relatedness to the Nicobarese, and to the other Austro-Asiatic speakers of mainland India and Southeast Asia. Based on the analysis of wide array of Y Chromosome SNP and STR markers among a large number of Austro-Asiatic tribes, representing entire micro-geographic and linguistic variation among them, Kumar et al. (2007) suggested that the Austro-Asiatic tribes of Andaman Nicobar Islands were probably the descendants of Mon-Khmer populations from Southeast

Asia, who in turn traced their origin from the Mundari populations of mainland India. However, the data suggest their colonization of Andaman and Nicobar Islands as a much later point of time. Using relatively ancient DNA retrieved from museum specimens mitochondrial DNA sequences of 11 Andaman Islanders were obtained by Endicott et al. (2003) and the mtDNA data suggest long-term isolation of the Andamanese, extensive population substructure, and/or two temporally distinct settlements. They conclude that the Negrito features of the Andaman Islanders are due to convergence rather than to common ancestry with Africans. Further they claim that their data support the Southern route hypothesis, though Cordaux and Stoneking (2003) differ with this conclusion and opine that the southern route hypothesis should await adequate genetic support (c.f. Tripathy et al. 2008).

Study by Anthropological Survey of India:

Mitochondrial DNA lineages sampled so far from south Asia, eastern Asia and Australasia show non-overlapping distributions of haplogroups within pan Eurasian M and N macrohaplogroups. Likewise, support from the archaeology is still ambiguous. However, mtDNA lineages sampled so far from South Asia, Eastern Asia and Australasia show non-overlapping distributions of haplogroups within pan Eurasian M and N macrohaplogroups. Likewise, support from the archaeology is still ambiguous. Keeping this in view, Kumar et al (2009) examined the reconstruction of Indian-Australian phylogenetic link. They sequenced 966-mitochondrial genomes from 26 relic tribes of India, and identified seven genomes, which share two synonymous polymorphisms with the M42 haplogroup, which is specific to Australian Aborigines. The results show a shared mtDNA lineage between Indians and Australian Aborigines provides direct genetic evidence of an early colonization of Australia through South Asia, following the "southern route".

To construct maternal phylogeny and prehistoric dispersals of modern human being in the Indian sub-continent, a diverse subset of 641 complete mitochondrial DNA (mtDNA) genomes belonging to macrohaplogroup M was chosen from a total collection of 2,783 control-region sequences, sampled from 26 selected tribal populations of India (Chandrasekhar et al. 2009). On the basis of complete mtDNA sequencing, They identified 12 new haplogroups - M53 to M64; redefined/ascertained and characterized haplogroups M2, M3, M4, M5, M6, M8′C′Z, M9, M10, M11, M12-G, D, M18, M30, M33, M35, M37, M38, M39, M40, M41, M43, M45 and M49, which were previously described by control and/or coding-region polymorphisms. The results indicate that the mtDNA lineages reported in the study (except East Asian lineages M8′C′Z, M9, M10, M11, M12-G, D) are restricted to Indian region. The deep-rooted lineages of macrohaplogroup ‘M’ suggest in-situ origin of these haplogroups in India. Most of these deep rooting lineages are represented by multiple ethnic/linguist groups of India. Hierarchical analysis of molecular variation (AMOVA) shows substantial subdivisions among the tribes of India (Fst = 0.16164). The current Indian mtDNA gene pool was shaped by the initial settlers and was galvanized by minor events of gene flow from the east and west to the restricted zones. Northeast Indian mtDNA pool harbors region specific lineages, other Indian lineages and East Asian lineages. They also suggest the establishment of an East.

The population genetics of the Indian subcontinent is central to understanding early human prehistory due to its strategic location on the proposed corridor of human movement from Africa to Australia during the late Pleistocene. Previous genetic research using mtDNA has emphasized the relative isolation of the late Pleistocene colonizers, and the physically isolated Andaman Island populations remain the source of claims supporting an early split between the populations that formed the patchy settlement pattern along the coast of the Indian Ocean. Using whole-genome sequencing, combined with multiplexed SNP typing, Barik et al (2008) investigated the deep structure of mtDNA haplogroups M31 and M32 in India and the Andaman Islands. The identification of a so far unnoticed rare polymorphism shared between these two lineages suggests that they are actually sister groups within a single haplogroup, M31′32. The enhanced resolution of M31 allows for the inference of a more recent colonization of the Andaman Islands than previously suggested, but cannot reject the very early peopling scenario. They further demonstrate a widespread overlap of mtDNA and cultural markers between the two major language groups of the Andaman archipelago. Given the "completeness" of the genealogy based on whole genome sequences, and the multiple scenarios for the peopling of the Andaman Islands sustained by this inferred genealogy, this study hints that further mtDNA based phylogeographic studies are unlikely to unequivocally support any one of these possibilities.

The high frequency of mitochondrial lineage "M2" consistent with its greater age and distribution suggests that it may represent the phylogenetic signature of earliest settlers. Accordingly, Kumar et al (2008) attempted to re-evaluate the impact and contribution of earliest settlers in shaping the genetic diversity and structure of contemporary Indian populations; using their newly sequenced 72 and 4 published complete mitochondrial genomes of this lineage. The M2 lineage, harbouring two deep rooting subclades M2a and M2b encompasses approximately one tenth of the mtDNA pool of studied tribes. The phylogeographic spread and diversity indices of M2 and its subclades among the tribes of different geographic regions and linguistic phyla were investigated in detail. Further, the reconstructed demographic history of M2 lineage as a surrogate of earliest settlers' component revealed that the demographic events with pronounced regional variations had played pivotal role in shaping the complex net of populations phylogenetic relationship in Indian subcontinent. Results suggest that tribes of southern and eastern region along with Dravidian and Austro-Asiatic speakers of central India are the modern representatives of earliest settlers of subcontinent. The Last Glacial Maximum aridity and post LGM population growth mechanised some sort of homogeneity and redistribution of earliest settlers' component in India. The demic diffusion of agriculture and associated technologies around 3 kyBP, which might have marginalized hunter-gatherer, is coincidental with the decline of earliest settlers' population during this period.

Chandrasekhar et al. (2007) scanned a total of 2169 samples from 21 tribal populations from different regions of India for the Y-chromosome Alu polymorphism. This study reports, for the first time, high frequencies (8-65%) of Y Alu polymorphic (YAP) insertion in Northeast Indian tribes. All seven Jarawa samples from the Andaman and Nicobar islands had the YAP insertion, in conformity with an earlier study of Andaman Islanders. One isolated case with haplotype E* was found in Dungri Bhill, a Western Indian population, while YAP insertion in Northeast India and Andaman tribes was found in association with haplotype D* (M168,

M174). YAP insertion frequencies reported in the mainland Indian populations are negligible, according to previous studies. Genetic drift may be the causative factor for the variable frequency of the YAP insertion in the mainland populations, while the founder effect may have resulted in the highest incidence of haplotype D among the Andaman Islanders. The results of YAP insertion and the evidence of previous mtDNA studies indicate an early out of Africa migration to the Andaman and Nicobar Islands. The findings of YAP insertion in Northeast Indian tribes are very significant for understanding the evolutionary history of the region (Chandrasekhar et al. 2007).

Anthropological Survey of India recently completed a study on "Genomic Diversity of 72 tribal Communities in India." Summary and conclusion of this study is as follows:

- Anthropological survey of India has studied communities comprising 7807 blood samples from different parts of the country under the project DNA Polymorphism in Contemporary Indian Population.

- Out of screened 5497 mtDNA genomes, Asian specific M lineage and European specific N lineage accounted for 68.2 % and 31.2% respectively. The frequency distribution of macrohaplogroup M and N varies significantly ($P < 0.02$) among studied populations with a cline from North (52% of M) towards Southern (77% of M) and Eastern (72% of M) regions of India.

- The Andamanese of Andaman Island, the Kattunayakan of Tamil Nadu, and the Chenchu of Andhra Pradesh do not possess European specific haplogroup N while they have 100 per cent haplogroup M.

- The Nicobarese of Nicobar island harbor hundred percent of haplogroup N and they do not exhibit Asian specific M haplogroup.

- A total of 3796 mtDNA genomes have been chosen for complete sequencing to construct maternal phylogeny and 2776 Y-Chromosomes to construct paternal Phylogeny of India.

- The present study identified 61 maternal lineages in Indian populations out of 81 world maternal lineages of Asian M haplogroup and all the major 14 Europen N lineages.

- The highest frequencies worldwide of macrohaplogroup M are observed in Asia, Australia, Siberia, Americas and East Africa.

- This M haplogroup has deep time depth >50,000 years of Western, Central, Southern and Eastern Indian haplogroups M2, M38, M54, M58, M33, M6, M61, M62 and do not rule out the possibility of macrohaplogroup M arising in Indian population.

- The present study confirmed Out-of-Africa hypothesis regarding origin, migration of Anatomically Modern Human. The ancestors of haplogroup M dispersed from Africa through

the southern route across the Horn of Africa along the coastal regions of Asia onwards to New Guinea and Australia.

- The results indicate that the mtDNA lineages reported in the present study (except East Asian lineages M8′C′Z, M9, M10, M11, M12-G, D) are restricted to Indian region. The deep rooted lineages of macrohaplogroup 'M' suggest in-situ origin of these haplogroups in India. Most of these deep rooting lineages are represented by multiple ethnic/linguist groups of India.

- Indian M haplogroup founder age is estimated as 66,000 years ± 9000 years. Hierarchical analysis of molecular variation (AMOVA) shows substantial subdivisions among the tribes of India (Fst = 0.18895). Molecular diversity is more with the population than among the population. The present populations are expanding after constrains in the past. The current Indian mtDNA gene pool was shaped by the initial settlers and was galvanized by minor events of gene flow from the East and West to the restricted zones.

- Northeast Indian mtDNA pool harbors region specific lineages, other Indian lineages and East Asian lineages. Indian gene pool also exhibits link with Andaman islanders through M31 haplogroup, with Australians through M42 haplogroups and with South East Asia Islanders through N22 haplogroup.

- One of the Andaman's specific haplogroup, M31 has been identified in Pauri Bhuiya, Munda of East India, Savara of Southeast part of India, Paite, Bohi Khasi of Northeast India. Previous genetic research using mtDNA has emphasized direct settlement of African exodus (Pleistocene colonizers) in the Andaman Island.

- But this study rules out the previous claim and unearthed that it is the Indians who migrated from India to Andaman Islands leaving their footprints in the mainland populations as mentioned above. The upper boundary for their migration has been calculated as ~24,000 years ago.

- Australian specific haplogroup M42 foot prints have been found in the Kutia Khond of Orissa, the Savara of Andhra Pradesh, Munda of Jharkhand, Madia of Maharastra and a sample of Pauribhuiya from Orissa. The coalescence time estimate is 55.2 ± 10.8 KYA (Thousand years ago) as divergence time of the Indian and Australian M42 coding-region sequences.

- For the first time it was identified mitochondrial DNA from one South Indian Savara individual that shares seven specific mutations with the N22 lineage observed in the Orang Asli group of Aboriginal Malaya, Cuyonin from Palawan and one single mutation sharing with Mindanao of Philippines which forms N22b sub-lineage. The coalescence time of N22 lineage is 20,600 ± 7000 thousand years before present. It shifts the focus of "Two layer" hypothesis of human settlement in ISEA (Island South East Asia).

- Indian Y chromosome phylogeny reveals that haplogroup H and O are the predominant haplogroups that constitute major original Indian gene pool. C5, F*, K*, P*, R1, R2 and L are the other haplogroups found among the tribes of India.

- Indian Y genome mainly comprises of 33 per cent of O haplogroup alleles, 22 per cent of H haplogroup alleles, followed by 16 percent of R alleles. This also report the presence of the J2 haplogroup alleles (3%) in tribal populations, where predicts its influx from the Fertile Crescent.

- The present study finds linguistic affiliation with Y lineages; Dravidian groups are predominantly constitutes H haplogroup reaching to 93% in the Koraga community. Among Indo-European linguistic groups of India, haoplogroup H reaches to 87% in the Nihal community. Haplogroup H is almost absent in the Tibeto-Burman populations of India while, sub haplogroup O2a is predominant among Austro-Asiatic groups whereas, O3a3c1 is predominant in Tibeto-Burman population groups.

- Thus major parental lineages of these seventy-one primitive groups constitute the original Indian gene pool. The minor lineages of India are the indicators of several waves of migrations into India. Thus the current Indian paternal gene pool was shaped by the initial settlers and was galvanized by the influx of genes from the west and east of India. Out of 20 major male lineages of the world, India possesses 15 male lineages.

- Thus deep rooting, high haplogroup diversity and antiquity of mtDNA and Y – Chromosome haplogroups reveals that Indian sub-continent acted as incubator for African exodus modern man at ~60-80,000 years ago, in peopling of world (Australia, East Asia, and South Asia, which is source for Central Asia).

- The generated mtDNA and Y-Chromosome data base will be an essential guide for diseases, Anthropological and forensic application in the Indian sub-continent.

Emerging Trend:

Studies based on mtDNA suggest deep-rooted maternal lineages for both caste and tribal populations of India and support single initial origin. The present population structure might have therefore resulted from subsequent admixture and drift. However, the estimated amount of west Eurasian or East Eurasian haplotype sharing of the different Indian populations has been vastly different among different studies. The mtDNA variation reported in India is largely explained as due to different level of admixture and genetic drift. Y-Chromosome sharing between Northeast India and South China is inferred as due to Neolithic expansion via Northeast Indian Corridor (Su et al. 2000; Basu et al. 2003), Kumar et al. (2007) on the other hand suggested migration of Austro-Asiatic males from India to Southeast Asia via Northeast Indian corridor, as the missing genetic link is evident in the form of Austro-Asiatic Khasi that were not genetically explored earlier. Autosomal markers have been used mostly for studying

the genetic affinities among different populations and are generally not used for estimating the time of common origin. The studies based on Autosomal markers have also provided some contrasting results. Some studies which focused on populations in a small geographical region found that genetic distance are not correlated with geographical distance (Viswanathan et al. 2004; Reddy et al. 2005), and others reported that they are correlated with geographical distance and/or ethnic affiliation (Kashyap et al. 2004b; Kritika et al. 2006; Bamshad et al. 2001; Reddy et al.; 2001a, 2001b), when populations were compared at broader level (Reddy et al. 2005; Langstieh et al. 2004; Majumder et al. 1999) the clusters were formed consistent to geographic affiliation of the groups. Roychoudhury et al. (2001), on the other hand, reported close affinity of the populations of similar linguistic background, Basu et al. (2003) found no tangible clustering based on social, geographic or linguistic criteria. Studies on Kivisild et al. (2003), Watkins et al. (2005) and Kashyap et al. (2006) on the contrary suggest common ancestry of Indian caste and tribal populations (c.f. Tripathy et al. 2008).

SUMMARY

One of the first studies based on mt DNA variation among the Indian population was attempted by Mountain et al (1995). They studied the three culturally divergent endogamous caste groups from coastal south western India. Their study indicates that Indian populations represent a major expansion possibly originated in Southern Asia. In a study of 250 individuals from 12 caste populations from northeastern Andhra Pradesh, Southern India, Bamshad et al. (1998) found the differences in social rank between castes correspond to mtDNA distances, suggesting genetic stratification corresponding to social stratification and this they interpreted as due to the system of hypergamy, wherein women of lower caste hierarchy is allowed to marry men of higher castes (c.f. Tripathy et al. 2008).

Indian population is closer to Caucasians and has an admixture with Asians. The North Indian population appears to have a recent admixture of the Caucasian mtDNA types which is absent in the south. This supports the recent peopling of the Indo-European language-speaking people in India (Barnabas et al. 1996). Prasad et al. (2001) of the opinion that dispersal of southern Chinese into mainland Southeast Asia may have included a westward expansion and colonization of the islands of the Andaman Sea. Thangaraj et al. (2005) found from the analysis of 3239 individuals of 58 endogamous populations of India that the frequency of 9bp deletion/insertion does not vary significantly among different and linguistic populations.

Mitochondrial DNA (mt DNA) variation of 23 ethnic populations of India drawn from diverse cultural, linguistic and geographical backgrounds is presented by Roychoudhury et al. (2000). There is extensive sharing of a small number of mtDNA haplotypes, reconstructed on the basis of restriction fragment length polymorphisms, among the populations. This shows that Indian populations were founded by a small number of females, possibly arriving on one of the early waves of out-of-Africa migration of modern humans. The authors found that the Asian-specific haplogroup M is in high frequency in most populations, especially tribal populations and Dravidian populations of Southern India. They also found that populations in which the frequencies of haplogroup M are relatively lower show higher frequencies of haplogroup U. These populations are primarily caste populations of northern India. This finding is indicative of a higher Caucasoid admixture in northern Indian populations. By examining the sharing of haplotypes between Indian and south-east Asian populations the

authors have provided evidence that south-east Asia was peopled by two waves of migration, one originating in India and the other originating in southern China. Terros et al. (2007) studies the mtDNA variation in the two Muslim sects from the Northern Indian province of Uttar Pradesh, the Sunni and Shia. A comparison of this data to that from Middle Eastern, Central Asian, North East African, and other Indian groups revealed that, at the mtDNA haplogroup level, both of these Indian-Sunni and Shia populations are more similar to each other and to the other Indian groups than to those from the other regions. These two Muslim sects exhibit a conspicuous absence of West Asian mtDNA haplogroups suggesting that their maternal lineages are of Indian origin (c.f. Tripathy et al. 2008).

In contrast to mtDNA, Y-Chromosome distances showed no correlation with social rank. This was interpreted as due to lack of male gene flow between castes of different social rank. Bamshad et al. (1998) concluded that the genetic stratification of the Hindu caste system is driven by the social mobility of women in a patrilocal society and social laws of hypergeny played a role in it. Linguistic evidence suggests that West Asia and Central Asia have been the two major geographical sources of genes in the contemporary Indian gene pool (Mukherjee et al. 2001). Are the Hindu castes largely derived from Indian local populations (i.e. tribal groups) or from recent immigrants to India? Archaeological and linguistic evidence support the latter hypothesis, whereas recent genetic data seem to favour the former hypothesis. Cordaux et al. (2004b) analyse the most extensive dataset of Indian caste and tribal Y chromosomes to date. They found that caste and tribal groups differ significantly in their haplogroup frequency distributions; caste groups are homogeneous for Y chromosome variation and more closely related to each other and to central Asian groups than to Indian tribal or any other Eurasian groups. The study conclude that paternal lineages of Indian caste groups are primarily descended from Indo-European speakers who migrated from central Asia approximately 3,500 years ago. Conversely, paternal lineages of tribal groups are predominantly derived from the original Indian gene pool.

The issue of the origin of Andaman and Nicobar tribes has always been interesting to Anthropologists. These tribes have been proposed to provide footprints of migration of early wave of modern humans to Australia (Redd et al. 2002; Macaulay 2005). Few studies have focused on the molecular genetic features of populations from Andaman and Nicobar islands. Analyzing mtDNA sequences and RFLP polymorphisms, Y-Chromosome biallelic marker and microsatellites in the populations of Andaman and Nicobar Islands, Thangaraj et al. (2003) observed low genetic variation among them. The closer genetic affinity of Andamanese to the Asians than to African populations suggests that they are the descendants of the early Paleolithic colonizers of South-East Asia. The results of Anthropological Survey of India indicate that the mtDNA lineages (except East Asian lineages M8′C′Z, M9, M10, M11, M12-G, D) are restricted to Indian region. The deep rooted lineages of macrohaplogroup ‘M’ suggest in-situ origin of these haplogroups in India. Most of these deep rooting lineages are represented by multiple ethnic/linguist groups of India.

ACKNOWLEDGEMENT

The author is grateful to Dr. Harshawaradhana, SA (P), Anthropological Survey of India, Dehradun for his guidance and inspiration for carrying out this work. Author is also grateful to an anonymous reviewer for valuable comments on the manuscript.

REFERENCES

Ashma R. Kashyap VK. 2003. Genetic profile based upon 15 microsatellites of four caste groups of the eastern Indian state, Bihar. *Ann Hum Biol*, 30: 570-578.

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.

Bamshad M. Fraley AE. Crawford MH. Cann RL. Busi BR. Naidu JM. Jorde LB. 1996. Mt DNA variation in Caste Populationas 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.

Bamshad M. Kivisild T. Watkins WS. Dixon ME. Ricker CE. Rao BB. Naidu JM. Prasad BVR. Reddy PG. Rasanayagam A. Papiha SS. Villems R. Reddy AJ. Hammer MF. Nguyen SV. Caroll ML. Batzer MA. Jorde LB. 2001. Genetic evidence on the Origins of Indian Caste Populations. *Genome Res,* 11: 994-1004.

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.

Barik SS. Sahani R. Prasad BVR et al. 2008. Detailed mtDNA Genotypes permit a Reassessment of the Settlement and population structure of Andaman Islands. *Am J Phys Anthropol*, 136:19-27.

Barua S. 2007. *Human Genetics*. Classique Books, Kolkata.

Basu A. Mukherjee N. Roy S. Sengupta S. Banerjee S. Chakraborty M. Dey B. Roy M. Roy B. Bhattacharyya NP. Roychouchudhury S. Majumder PP. 2003. Ethnic India: A Genomic View, With Special Reference to Peopling and Structure. *Genome Res*, 13: 2277-2290.

Bhattacharyya NP. Banerjee R. Roy B. Roychoudhury S. Majumder PP. 1999. Negligible Male Gene Flow Across ethnic Boundaries in India, Revealed by Analysis of Y-Chromosomal DNA Polymorphism. *Genome Res*, 9: 711-719.

Chandrasekhar A. Kumar. Sreenath J. Sarkar BN. Urade BP. Mallick S. Bandyopadhyay SS. Barua P. Barik SS. Basu D. Kiran U. Gangopadhyay P et al. 2009. Updating Phylogeny of Mitochondrial DNA macrohaplogroup M in India: Dispersal of Modern Human in South Asian Corridor. *Plos One*, 4(10)e 7447.

Chandrasekhar A. Saheb SY. Gangopadhyay P et al. 2007. YAP insertion signature in South Asia. *Annals of Human Biology*, 34(5):582-586.

Clark VJ. Sivendren S. Saha N. Bentley GR. Aunger R. Sirajuddin SM. Stoneking M. 2000. The 9-bp deletion between the Mitochondrial Lysine tRNA and COII Genes in Tribal populattions of India. *Hum Biol*, 72: 273-285.

Cordaux R. Stoneking M. 2003. South Asia, the Andamanese and the genetic evidence for an "early" human dispersal out of Africa. *Am J Hum Genet*, 72: 1586-1590.

Cordaux R. Saha N. Bentley GR. Anger R. Sirajuddin SM. Stoneking M. 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur J Hum Genet*, 11: 253-264.

Cordaux R. Weiss G. Saha N. Stoneking M. 2004a. The northeast Indian passageway: a barrier on corridor for human migrations? *Mol Biol Evol*, 21: 1525-1533.

Cordaux R. Aunger R. Bentley G. Nasidze I. Sirajuddin SM. Stoneking M. 2004b. Independent origins of Indian caste and tribal paternal lineages. *Curr Biol*, 14: 231-235.

Dutta R. Reddy BM. Chattopadhyay P. Kashyap VK. Sun G. Deka R. 2002. Patterns of genetic diversity at the nine forensically approved STR loci in the Indian populations. *Hum Biol*, 74: 33-49.

Ember CR. Ember M. Peregrine PN. 2003. *Anthropology.* Tenth Edition. Pearson Education, Delhi.

Endicott P. Gilbert MT. Stringer C. Lalueza-Fox C. Willerslev E. Hansen AJ. Cooper A. 2003. The Genetic origins of the Andaman Islanders. *Am J Hum Genet*, 72: 178-184.

Gaikwad S. Vasulu TS. Kashyap VK. 2006. Microsatellite diversity reveals the interplay of language and geography in shaping genetic differentiation of diverse Proto-Australoid populations of west-central India. *Am J Phys Anthropol*, 129: 260-267.

Genomic Diversity of 72 tribal Communities in India. Anthropological Survey of India, Unpublished.

Ghoshmaulik SK. 2000. *Classification and Origin of People of India: An Unsolved Problem.* Sectional President Address. 87th Session of Indian Science Congress Association.

Gutala R. Carvalho-Silva DR. Jin L. Yngvadottir B. Avadhanula V. Nanne K. Singh L. Chakraborty R. Tayler-Smith C. 2006. A shared Y-chromosomal heritage between Muslims and Hindus in India. *Hum Genet*, 120: 543-551.

Kashyap VK. Ashma R. Gaikwad S. Sarkar BN. Trivedi R. 2004b. Deciphering diversity in populations of various linguistic and ethnic affiliations of different geographical regions of India: analysis based on 15 microsatellite markers. *J Genet*, 83: 49-63.

Kashyap VK. Guha S. Sitalaximi T. Bindu GH. Hasmain SE. Trivedi R. 2006. Genetic structure of Indian population based on fifteen autosomal microsatellite loci. *BMC Genet*, 7: e28.

Kashyap VK. Chattopadhyay P. Dutta R. Vasulu TS. 2004a. Genetic structure and affinity among eight ethnic populations of Eastern India:based on 22 polymorphic DNA loci. *Am J Hum Biol*, 16: 311-327.

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. 1999. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol*, 9: 1331-1334.

Kivisild T. Rootsi S. Metspalu M. Mastana S. Kaldma K. Parik J. Metspalu E. Adojaan M. Tolk HV. Stepanov V. Golge M. Usanga E. Papiha SS. Cinnioolu C. King R. Cavalli-Sfroza L. Underhill PA. Villems R. 2003. The generic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet*, 72: 313-332.

Krithika S. Trivedi R. Kashyap VK. Bharati P. Vasulu TS. 2006. Antiquity, geographic contiguity and genetic affinity among Tibeto-Burman populations of India: A microsatellite study. *Ann Hum Biol*, 33: 26-42.

Kumar S. Ravuri RR. Koneru P. Urade BP. Sarkar BN. Chandrasekhar A. Rao VR 2009. Reconstructing Indian-Australian Phylogenetic link. *BMC Evolutionary Biology,* 9:193.

Kumar S. Padmanabham PBSV. Ravuri RR et al. 2008. The earliest settlers antiquity and evolutionary history of Indian Populations: evidence from M2 mtDNA lineage. *BMC Evolutionary Biology,* 8:230; 1471-2148-8:230.

Kumar V. Langstieh BT. Madhavi KV. Naidu VM. Singh HP. Biswas S. Thangraj K. Singh L. Reddy BM. 2006b. Global patterns in Human Mitochondrial DNA and Y-Chromosome variation caused by Spatial instability of the Local Cultural Processes. *Plos Genet,* 2 (4): e53.

Kumar V. Langsiteh BT. Biswas S. Babu JP. Rao TN. Thangaraj K. Reddy AG. Singh L. Reddy BM. 2006. Asian and non-Asian origins of Mon-Khmer- and Mundari-Speaking Austro-Asiatic Populations of India. *Am J Hum Biol*, 18: 461-469.

Kumar V. Reddy AN. Babu JP. Rao TN. Langsteieh BT. Thangraj K. Reddy AG. Singh L. Reddy BM. 2007. Y- chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evol Biol*, 7: 47.

Langstieh BT. Reddy BM. Thangaraj K. Kumar V. Singh L. 2004. Genetic diversity and relationships among the tribes of Meghalaya compared to other Indian and Continental populations. *Hum Biol,* 76: 569-590.

Langstieh BT. 2007. *Human Genomic Studies in India- A Review on Anthropo-Genetics.* pp. 36-49. In: IJS Bansal and A. Basu (eds.), Some Dimensions of Human Biological Diversity Among Indian Populations. INCAA Occasional Papers. INCAA, Kolkata.

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 H. Oppenheimer S. Torroni A. Richards M. 2005. Single, Rapid Coastal settlement of Asia Revealed by Analysis of Complete Mitochondrial Genome. *Science,* 308: 1034-1036.

Majumder PP. Roy B. Banerjee S. Chakraborty M. Dey B. Mukherjee N. Roy M. Thakurta PG. Sil SK. 1999. Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *Eur J Hum Genet,* 7: 435-446.

Majumder PP. 2001. Ethnic populations of India as seen from an evolutionary perspective. *J Biosci,* 26:533-545.

Metspalu M. Kivisild T. Metspalu E. Parik J. Hudjasov 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 southwestern Asia were likely shaped during the initial settlement of Eurasia by automatically modern humans. *BMC Genet*, 5: 26

Mountain JL. Hebert JM. Bhattacharya S. Underhill PA. Ottolenghi C. Gadgil M. Cavalli-Sforza LL. 1995. Demogrphic History of India and mtDNA-sequence diversity. *Am J Hum Jenet*, 56:979-992.

Mukherjee N. Nebel A. Oppenheim A. Majumdar PP. 2001. High-resolution analysis of Y-chromosomal polymorphisms reveals signatures of population movements from Central Asia and West Asia into India. *J Genet*, 80:125-135.

Palanichamy MG. Sun C. Agarwal S. Bandelt HJ. Kong QP. Khan F. Wang CY. Chaudhuri TK. Palla V. Zhang YP. 2004. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am J Hum Genet*, 75: 966-978.

Passarino G. Semino O. Bernini LF. Santachiara-Bnerecetti 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 Caucasoid; it probably predates the diversification between Indians and Orientals. *Ann Hum Biol*, 23:121-126.

Prasad BV. Ricker CE. Watkins WS. Dixon ME. Rao BB. Naidu JM. Jorde LB. Bamshad M. 2001. Mitochondrial DNA variation in Nicobarese Islanders. *Hum Biol*, 73: 715-725.

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

Rajkumar R. Kashyap VK. 2004. Genetic structure of four socio-culturally diversified caste populations of southwest India and their affinity with related Indian and global groups. *BMC Genet*, 5:23.

Ramana GV. Su B. Jin L. Singh L. Wang N. Underhill P. Chakraborty R. 2001. Y-Chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe and the migrant Siddi populations of Andhra Pradesh, South India. *Eur J Hum Genet*, 9: 695-700.

Redd AJ. Roberts-Thomson J. Karafet T. Bamshad M. Jorde LB. Naidu JM. Walsh B. Hammer MF. 2002. Gene flow from the Indian subcontinent to Australia: evidence from the Y chromosome. *Curr Biol*, 12: 673-677.

Reddy BM. Naidu VM. Madhavi VK. Thangaraj LK. Kumar V. Langstieh BT. Venkataramana P. Reddy AG. Singh L. 2005. Microsatelite diversity in Andhra Pradesh, India: genetic stratification versus social stratification. *Hum Biol*, 77:803-823.

Reddy BM. Sun G. Luis JR. Crawford MH. Hemam NS. Deka R. 2001a. Genomic diversity at Thirteen Short Tandem Repeat Loci in a Substructured Caste Population, Golla, of Southern Andhra Pradesh, India. *Hum Biol*, 73: 175-190.

Reddy BM. Dutta R. Langsteih BT. Kashyap VK. 2001b. Diversity at three tetrameric STR loci in a substructured Golla caste population of Southern Andhra Pradesh, in comparison to other Indian populations. *Int J Hum Genet*, 1: 101-108.

Reddy BM. Langstieh BT. Kumar V. Nagaraja T. Reddy ANS. Meka A. Reddy AG. Thangaraj K. Singh L. 2007. Austro-Asiatic Tribes of Northeast India Provide Hitherto Missing Genetic Link between South and Southeast Asia. *PLoS ONE*, 2(11):e1141.

Roychoudhury S. Roy S. Dey B. Chakraborty M. Roy M. Roy B. Ramesh A. Prabhakaran N. Rani MVU. Viswanathan H. Mitra M. Sil SK. Majumder PP. 2000. Fundamental Genomic unity of ethnic India is revealed by analysis of mitochondrial DNA. *Current Science*, 79: 1182-1192.

Roychoudhury S. Roy S. Basu A. Banerjee R. Viswanathan H. Usha Rani MV. Sil SK. Mitra M. Majumder PP. 2001. Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum Genet*, 109: 339-350.

Sahoo S. Kashyap VK. 2005. Influence of language and ancestry on genetic structure of contiguous populations: a microsatellite based study on populations of Orissa. *BMC Genet*, 6(1):4.

Sahoo S. Kashyap VK. 2006. Phylogeography of mitochondrial DNA and Y-chromosome haplo-groups reveal asymmetric gene flow in populations of Eastern India. *Am J Phys Anthropol*, 131: 84-97.

Sengupta S. Zhivotovsky LA. King R. Mehdi SQ. Edmonds CA. Chow CT. Lin AA. Mitra M. Sil SK. Ramesh A. Usha Rani M.V. Thakur CM. Cavalli-Sforza LL. Majumder PP. Underhill P. 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. Amm J *Hum Genet*, 78(2):202-221.

Su B. Xiao C. Deka R. Seielstad MT. Kangwanpong D. Xiao J. Lu D. Underhill P. Cavalli-sforza L. Chakraborty R. Jin L. 2000. Y-Chromosome haplotypes reveal pre-historical migrations to the Himalayas. *Hum Genet*, 107:582-590.

Sun C. Kong QP. Palanichamy MG. Agarwal S. Bandelt HJ. Yao YG. Khan F. Zhu CL. Chaudhuri TK. Zhang YP. 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol Biol Evol*, 23: 683-690.

Terros MC. Rowold D. Luis JR. Khan F. Agarwal S. Herrera RJ. 2007. North Indian Muslims: Enclaves of foreign DNA or Hindu converts? *Am J Phys Anthrop*, 133: 1004-1012.

Thangaraj K. Singh L. Reddy AG. Rao VR. Sehgal SC. Underhill PA. Pierson M. Frame IG. Hagelberg E. 2003. Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr Biol*, 13: 86-93.

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

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

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

Tripathy V. Nirmala A. Reddy BM. 2008. Trends in Molecular Anthropological Studies in India. *Int J Hum Genet*, 8 (1-2):1-20.

Trivedi R. Sitalaxmi T. Banerjee J. Singh A. Sircar PK. Kashyap VK. 2006. Molecular insights into the origins of the Shompen, a declining population of the Nicobar archipelago. *J Hum Genet*, 51: 217-226.

Viswanathan H. Deepa E. Cordaux R. Stoneking M. Rani MVU. Majumder PP. 2004. Genetic structure and affinities among tribal populations of southern India: a study of 24 autosomal DNA markers. *Ann Hum Genet*, 68: 85-177.

Watkins WS. Bamshad M. Dixon ME. Bhaskara Rao B. Naidu JM. Reddy PG. Prasad BV. Das PK. Reddy PC. Gai PB. Bhanu A. Kusuma YS. Lum JK. Fischer P. Jorde LB. 1999. Multiple origins of the mtDNA 9-bp deletion in populations of South India. *Am J Phys Anthropol*, 109: 147-158.

Watkins WS. Prasad BV. Naidu JM. Rao BB. Bhanu BA. Ramchandran B. Das PK. Gai PB. Reddy PC. Reddy PG. Sethuraman M. Bamshad MJ. Jorde LB 2005. Diversity and divergence among the tribal populations of India. *Ann Hum Genet*, 69: 680-692.