

Trends in Molecular Anthropological Studies in India

Vikal Tripathy¹, A. Nirmala² and B. Mohan Reddy¹

1. *Molecular Anthropology group, Biological Anthropology Unit, Indian Statistical Institute, Hyderabad 500 007, Andhra Pradesh, India,*

2. *School of Human Genome Research and Genetic Disorders, Mahathma Gandhi National Institute of Research and Social Action (MGNIRSA), Hyderabad 500 029, Andhra Pradesh, India*

KEYWORDS Population structure; molecular anthropology; mtDNA; Y-chromosome; autosomal DNA; India

ABSTRACT Indian population is characterized by wide diversity and unique population structure shaped by different waves of migration and the practice of caste endogamy. Anthropologists have been studying the peopling of India and the relationships between different populations using traditional genetic markers. With the advent of molecular genetic techniques the focus has turned to using the DNA polymorphisms for resolving different anthropological questions and to test the different hypotheses in vogue. In this paper we make an attempt to critically review the trends in molecular anthropological studies till date and bring out salient features of the findings. An attempt has been made to evaluate the merit of the molecular studies in the perspective of unique population structure of India.

INTRODUCTION

Anthropology is defined as a study of man in time and space. In simple words the subject matter of Anthropology deals with Human variation and evolution, humans not as individuals but ethnic groups or populations as a units of study. Although this subject is broadly divided into Cultural/Social and Biological/Physical Anthropology, anthropologists ideally, study man holistically, taking into account both the biological and socio-cultural aspects of man. The subject had its beginning in the studies of small marginalized tribal communities but today any human population comes under its ambit.

Biological anthropology started of as descriptive science of physical variation of humans of both present and past (fossils). One major aim of studying this variation was to study human evolution apart from the inherent desire of knowing the "other". In the beginning, anthropometric and anthroposcopic techniques were developed and thoroughly used by anthropologist to study human variation, to describe humans through out the world and to classify them

into different races. Although dermatoglyphics were also used along with anthropometry in describing ethnic and/or population variation, it was rarely used in the racial classification. The discovery of genetically determined blood cell polymorphism in the beginning of twentieth century and later on protein polymorphism provided anthropologist with new tools to study human variation. Anthropometry, dermatoglyphics and serological markers were used extensively in the twentieth century for studying the evolutionary relationships between different populations both at the regional and global levels. The problem with anthropometric variables had been the environmental noise as they are influenced by many confounding variables besides being genetically multifactorial. The classical genetic markers, which were mostly serological, were used extensively to investigate population relationships across the world, both at micro and macro levels. The role of different evolutionary forces like drift, selection, mutation, migration and gene flow in shaping the genetic make up of populations has also been investigated on the basis of classical genetic markers. Cavalli-Sforza et al. (1994) attempted a synthesis of the classical genetic marker data available on populations from different parts of the world. One drawback of the classical markers in the study of evolutionary history and population relationships has been that one cannot be sure under what selective pressures

Corresponding Author's Address: Professor B. Mohan Reddy, Ph.D. Molecular Anthropology Group, Biological Anthropology Unit, Indian Statistical Institute
Street No. 8, Habsiguda, Hyderabad 500 007,
Andhra Pradesh, India
Telephone: 040-27171906; Extn: 27;
E-mail: bmr@isical.ac.in (OR) bmrishi@gmail.com

the genetic markers have been under different environmental stresses. More over, few of the classical genetic markers taken as one trait could have been a group of different polymorphism at the DNA level (for example, G6PD deficiency; Tripathy and Reddy 2007).

The advent of PCR techniques in the 1980s along with the use of restriction enzymes and later the use of sequencing methods for identifying polymorphisms at the DNA level provided anthropologist with new and more powerful techniques and markers for testing different anthropological hypotheses. The study of Cann et al. (1987), which gave support to the Out of Africa theory of human evolution, revolutionized the field of molecular anthropology and within a decade anthropologists/population geneticists started using DNA sequence data for understanding human evolution and for characterizing human populations based on Haplotype and Haplogroup frequencies.

We present here a review of the Molecular Anthropological studies carried out on Indian populations in the background of the unique population structure of India. Although the scope of molecular anthropology can extend to population based approaches in molecular epidemiology and adaptation such studies are very few and far apart or are lacking on Indian populations. Therefore, the focus of this review pertains basically to the molecular genetic studies on population structure and peopling of India.

THE UNIQUE INDIAN POPULATION STRUCTURE

Indian subcontinent is known for its enormous diversity, cultural as well as biological. This diversity owes itself to the innumerable waves of migration from different parts of the world at different points of time. These migrants brought with them not only languages and cultures but also genes. As a result, the present day Indian population is a conglomeration of diverse ethnic elements, language families and cultures. In India, one finds almost all the ethnic constituents that can be found anywhere in the world as there are populations with physical features of Australoids, Caucasoids, Mongoloids, and Negroids. Indian populations speak innumerable languages that can be broadly grouped into Austro-Asiatic, Dravidian, Indo-European and Tibeto-Burman linguistic families. However, the most unique

feature of Indian population structure is the division of its population, within each linguistic area, into strictly defined hierarchical endogamous castes, tribes and religious communities whose marriage, hence genetic boundaries are strictly impermeable. While languages form barriers between any two linguistic regions even between the same castes, within a linguistic region, each caste/tribe/religious group is further subdivided into subcastes/subtribes/subgroups, depending on the size, nature of distribution etc., besides the possibility of forming geographic/breeding isolates when a caste/subcaste or tribe is large and distributed in a wide geographic area. Because of the pattern of substructuring Indian population contains a large number of endogamous groups with isolated gene pools which have evolved for over 3000 years. The total number of endogamous groups is estimated to be around 40000 comprising about 37000 castes and subcastes and 3000 tribal, religious and other historical migrants populations (Gadgil and Malhotra 1983; Malhotra 1984). On the other hand, Anthropological Survey of India (Singh 1993) identified 4635 communities which are, strictly speaking, not endogamous units. The cultural patterns governing marriage vary between the Dravidian and Indo-European kinship system (Table 1) which is expected to result in high degree of inbreeding and much smaller effective population sizes of the populations in the south as compared to those in the northern parts of India. Overall, the Indian populations show extreme variation in size, nature of sub-structuring etc, hence in the rate of their microevolution.

Theories on the Peopling of India

With the wide acceptance of Out of Africa hypothesis, India is considered a major corridor of human evolution and expansion. Of the various groups which inhabit India the Austro-Asiatic speaking tribal groups are considered the first inhabitants of Indian subcontinent (Kumar and Reddy 2003, Thapar 1966, Risley 1915, Rapson 1955, Pattanayak 1998). Some like Buxton (1925) and Sarkar (1958) has supported Dravidians as the original immigrants.

The Skeletal and Archaeological evidences from India points towards an early habitation of humans on the Indian subcontinent and stone tool evidences have been found from early Stone Age. However, which species of genus Homo had

Table 1: Difference in marriage patterns in Dravidian and Indo-European kinship system

| <i>Dravidian</i> | <i>Indo European</i> |
|---|---|
| 1. Consanguineous marriages highly preferred | 1. Sapinda rule prohibits consanguineous marriages |
| 2. Village endogamy preferred | 2. Village exogamy is the norm |
| 3. No restriction of marriage with neighboring villages | 2. Marriages with neighboring villages not favored |
| 4. Limited marital network; restricted choice of mates | 3. Much wider marital network and greater choice of mates |

used these tools is not clear. Skeletal evidence of earliest Anatomical Modern Humans (AMH) remains is dated at around 8000 ybp (Malhotra 1978). The skeletal evidence points towards presence of both Australoid and Caucasoid elements at all the periods. The Mongoloid features are missing from the skeletal remains.

Conclusions from the Earlier Studies Based on Traditional Variables

A number of studies have reported gene frequencies of one or more traditional genetic markers on many Indian populations. Gene frequencies for different markers from different studies on Indian populations have been compiled and presented by Bhasin et al. (1992). They also attempted to find some patterns in average gene frequencies of groups of populations defined by geography, language, ethnicity and occupation (Bhasin et al. 1994; Bhasin and Walter, 2001). A few studies have attempted studying different populations of India using the genetic and anthropometric markers at regional and local level. Studies at the local level have examined the degree of biological similarity between endogamous groups living in a very restricted area or between subdivisions of a single caste/tribal population (Reid 1984). Indian population structure as revealed from the studies based on anthropometric, serological and other classical markers as summarized by Malhotra and Vasulu (1993) is presented below:

1. There is wide range of gene frequency in all genetic systems and a great deal of variation, sometimes, even in the same geographical region
2. With exception to Africa, India harbors more genetic diversity than the other comparable global regions.
3. Tribals are genetically and morphologically different from non-tribal populations. The southern Indian tribal populations are different from the central and north eastern tribal populations

4. Geographically contiguous populations are genetically more similar than those with linguistic affiliation.
5. The amount of genetic diversity in Indian Populations is comparable to that existing in the major races of the man (Majumder and Mukherjee 1993)

The classical genetic markers have also been used to study the relationships between different castes as well as between caste and tribal groups. The general conclusions which emerged from these studies are:

1. Large genetic diversity between castes belonging to two different Varnas.
2. Geographically contiguous castes are genetically and morphologically closer, irrespective of caste or social hierarchy
3. Genetic diversity among castes is 1-3%, whereas between castes and tribes it is about 5%.
4. Geographic clines are observed in some traits, viz. ABO, Sickle cell, HbE, G6PD, which indicates role of selection as well.
5. Processes of drift, founder effect, gene flow and the fission and fusion are responsible for the micro-evolutionary differentiations in Indian populations.
6. Average heterozygosity, more or less, is of same magnitude in tribal or caste populations.
7. Genetic differentiation between castes and tribes, measured by the ratio of genetic distance to the mean of the average heterozygosity is of the same order (0.006) as compared to that among the human races (Roychoudhury 1984).

Molecular Anthropology Studies

With this background of pre-molecular genetic studies and inferences on Indian population structure, we review the Molecular Anthropology studies on the Indian populations focusing on genetic variation and peopling of India. We pay specific attention to reviewing studies that tried to test different anthropological/sociological

hypotheses concerning population structure and peopling of India, routes of migration of modern Humans through India, the genetic basis of the caste system, and relationships among different caste and tribal populations.

The molecular anthropology studies can be divided into three categories based on the type of markers used:

- (i) *Mitochondrial DNA (mtDNA) Variation:* Maternally inherited, highly polymorphic. Inferences provide clues to maternal lineages.
- (ii) *Y Chromosome Variation:* Uniparental transmission along the male lines, small effective population size and absence of recombination (except pseudo-autosomal region). Suitable for tracing male initiated migrations.
- (iii) Autosomal DNA Variation

mtDNA VARIATION IN INDIA

Studies Based on Hypervariable Region I & II Sequences in mtDNA

One of the first studies reporting use of mtDNA variation among Indian populations was that of Mountain et al. (1995). mtDNA sequences obtained from three culturally divergent Indian endogamous caste groups from coastal south western India indicated that the Indian populations represented a major expansion possibly originating in Southern Asia. The date of expansion estimated at some time point after modern human initially left Africa. Bamshad et al. (1996) reported mtDNA variation in four caste populations, viz. Brahmin (9), Yadava (10), Kapu (7) and Relli (10) of Andhra Pradesh and compared them with African, European and East Asian populations. mtDNA diversity in Indian Caste populations was found to be intermediate between African and other continental populations. mtDNA variation in the Indian caste populations was more than that of Europeans and East Asians. Higher diversity among Indian populations next only to Africans had earlier been observed for classical genetic markers as well. In a study of 250 individuals from 12 Telegu-speaking caste populations from northeastern Andhra Pradesh, southern India, Bamshad et al. (1998) showed that differences in social rank between castes correspond to mitochondrial DNA (mtDNA) distances between castes, suggesting genetic stratification corre-

sponding 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.

RFLP Based Bi-allelic Marker Studies

Passarino et al. (1996a) used mtDNA haplotypes for studying variation in 70 individuals from Punjab and 96 from UP (Uttar Pradesh) and AP (Andhra Pradesh) and concluded that Indo-European migration affected the population structure mainly in the north and in the center of the Indian sub-continent, not in the south. Using 6 restriction enzymes, Barnabas et al. (1996) studied mtDNA variation of 100 Indians belonging to 14 different languages and found that the Indian population is closer to Caucasians and has an admixture with Asian. These data further suggested that the North Indian populations appears to have a recent admixture of the Caucasian mtDNA, which is absent in the South, thus supporting the recent peopling of the Indo-European language speaking people in India. On the other hand, Passarino et al. (1996b) found 50% of the Indians (in a sample of 133 Hindus and 30 tribals) to represent the two mtDNA haplotypes that were considered to be the ancient East Asian markers and not found among the Caucasians of the Mediterranean basin and concluded that the Indo-European speakers were genetically different from the pre-existing Indian population. They further speculated that the Indians behave as typical Caucasoid for other genes (other than mtDNA) because the relative contribution of the Indo-European speaker to final Indian genetic make-up was mostly a paternal one.

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

mtDNA 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 (COII) 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.

Majumder (2001) reported absence of 9bp

deletion in Indian populations as reflected from a study of some 30 population from different regions. On the other hand, Watkins et al. (1999), Clark et al. (2000), Prasad et al. (2001), Thangaraj et al. (2005) and Kumar et al. (2006a) reported that the 9bp deletion is present in some caste and tribal populations with varying frequencies. Watkins et al. (1999) studied 9bp deletion in mtDNA in 646 Indians from 12 caste and 14 tribals from south India and suggested that 9bp deletion has arisen independently in some Indian tribal populations. Other 9bp deletion haplotypes are of Asian and African origin. Different affinities for 9bp deletion among caste and tribal populations thus suggested different origins for these populations. Clark et al. (2000) analyzed 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 in 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.

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

Kumar et al. (2006a) analyzed 1,686 samples from 31 tribal populations of India for the mitochondrial DNA 9-bp deletion/insertion polymorphism. The results suggest multiple origins/migrations of Austro-Asiatic groups into the Indian Subcontinent and also distinct origin for Austro-Asiatic linguistic groups. Austro-Asiatic populations of India have come in multiple waves of migration, and the ancestors of present day Mundari groups might have been the first to

arrive in India through the western Indian corridor, subsequently migrating to Southeast Asia.

Haplogroup Based Studies

The study of Kivisild et al. (1999) for the first time contradicted the commonly accepted hypothesis of massive Indo-Aryan invasion some 3500 to 4000 years ago and thus recent massive admixture of Caucasoids with Indian populations. The genetic link for the common mtDNA haplogroup U, which roughly accounts for a fifth of mtDNA lineages of both these populations, was found to be very deep and dated to late Pleistocene. The coalescence time (~50,000yr) of the Indian-specific subset of the west Eurasian haplotype (U2i) suggests that the west Eurasian admixture may have been much older than the proposed Dravidian and Indo-European incursions based on language phylogeny. Estimate for this split for haplotype U2i is close to the estimated time for the peopling of Asia and the first expansion of anatomically modern humans in Eurasia and perhaps predates their spread to Europe. Only a small fraction of the 'Caucasoid-specific' mtDNA lineages found in Indian populations can be ascribed to a relatively recent admixture. Thus these results questioned the commonly held hypothesis of massive Indo-Aryan invasion to India some 4,000 years ago. Most subsequent studies yielded results confirmatory to the findings of Kivisild et al. (1999).

Roychoudhury et al. (2000) studied mtDNA variation among 23 ethnic populations of India from diverse cultural, linguistic and geographic backgrounds. They found extensive sharing of one or two haplotypes across population groups within India, irrespective of the geography, linguistic affinity or social proximity. Most of the mtDNA diversity observed in Indian populations was found between individuals within population. No significant structure of haplotype diversity by socio-religious affiliation, geography or linguistic affiliation was found, suggesting a fundamental unity of mtDNA lineages in India, in spite of the extensive cultural and linguistic diversity.

Roychoudhury et al. (2001) sampled tribal populations belonging to all the three language groups—Austro-Asiatic speakers: Santal (n=21), Munda (n=7), Lodha (n=32); Dravidian speakers:

Muria (n=49), Kota (n=45), Kurumba (n=54), Irula (n=50); Tibeto-Burman speakers: Tipperah (n=51). MtDNA RSP haplotypes showed 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. Analyses of haplogroup and HVS-1 sequence data provided evidence in support of the hypothesis that the Austro-Asiatic speakers are the most ancient inhabitants of India. Subsequently, this group analyzed data for mtDNA RSPs and HVS1 sequences of 44 populations (Basu et al. 2003) and found a significant sharing both in the number of sequences and in proportion of individuals sharing these sequences and suggested uniformity of female lineages across India. They also found that Austro-Asiatic tribals possess highest frequency of the ancient East Asian mtDNA haplogroup M and exhibit highest HVS1 nucleotide diversity, and used this evidence to reiterate support to the hypothesis that Austro-Asiatic speakers may be the earliest inhabitants of India. However, this conclusion was based on just 3 of the 30 Austro-Asiatic tribes inhabiting the Indian subcontinent. Given that the analyses of mtDNA was at the level of macro haplogroup, M, which is ubiquitous to Asia and other regions and due to inadequate representation of populations of different linguistic groups (particularly of Austro-Asiatics) with only a few samples these conclusions are at the best tentative and should await further evidence with formidable samples and better representation of the different linguistic groups.

Two tribal groups from southern India - the Chenchus and Koyas - were analyzed by Kivisliid et al. (2003) for variation in mitochondrial DNA (mtDNA), the Y chromosome, and one autosomal locus and were compared with six caste groups from different parts of India, as well as with western and central Asians. Variation in the mtDNA, Y-Chromosome and autosomes suggested 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 phylogeography of the primal mtDNA and Y-chromosome founders suggests that these southern Asian Pleistocene coastal settlers from Africa would have provided the inocula for the subsequent differentiation of the distinctive eastern and western Eurasian gene pools.

Bamshad et al. (2001) analysed mtDNA (HVS 1) and 14 RSPs, Y-Chromosome (5 STR and 20 SNPs) and autosomal (1 LINE-I and 39 Alu inserts) variation in ~265 males from 8 different Telugu speaking caste populations of AP in South India. Comparisons were made with ~400 individuals from tribal and Hindi speaking populations. For maternally inherited mtDNA, each caste was found to be more similar to Asians. However, 20%–30% of Indian mtDNA haplotypes belong to West Eurasian haplogroups, and the frequency of these haplogroups is proportional to caste rank, hence the affinity of different populations to Europeans is proportionate to caste rank. Upper castes have greater genetic affinity to Europeans than to Asians and the upper castes are also significantly more similar to Europeans than are the lower castes. Indian castes are more likely to be of proto-Asian origin with west Eurasian admixture resulting in rank related sex-specific differences in the genetic affinities of castes to Asians and Europeans. They conclude that Indo-European languages may not reflect a common origin of Europeans and most Indians, but rather underscores the transfer of languages mediated by contact between west Eurasians and native proto-Indians.

Cordaux et al. (2003) analyzed 370 bp of the first hypervariable region of the mitochondrial DNA (mtDNA) control region in 752 individuals from 17 tribal and four non-tribal groups from the Indian subcontinent. Southern Indian tribes showed reduced diversity and large genetic distances. By contrast, northern groups exhibited more diversity. Phylogenetic analyses revealed that southern and northern groups (except north-eastern ones) have related mtDNA sequences albeit at different frequencies. Within India, northeastern tribes are quite distinct from other groups. The Indian mtDNA gene pool appears to be more closely related to the east Eurasian gene pool than the west Eurasian one. Overall, analyses of molecular variance suggested that caste and tribal groups are genetically similar with respect to mtDNA variation.

Baig et al. (2004) analyzed mtDNA variation among seven communities belonging to tribes and castes of different hierarchy from western Maharashtra to test the hypothesis of tribal origin of some caste groups. Nucleotide diversity, gene diversity and average mismatches were found to of the same magnitude. The mtDNA haplogroups showed that both caste and tribal populations

share similar branches of the trees and the maternal lineages have their roots in early late Pleistocene. Thanseem et al. (2006) also 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.

Quintana-Murci (2004) studied the mtDNA variation in the southwestern and Central Asian regions, the area which has witnessed numerous waves of migration in the history of humankind. The 910 mitochondrial DNAs (mtDNAs) he analyzed represented 23 populations of Iranian plateau, Indus Valley and Central Asia. Indus Valley population comprised of populations from Pakistan and a group of Gujaratis from India. Both the geographical distribution of lineages and spatial analysis of molecular variance showed that populations located west of the Indus Valley mainly harbor mtDNAs 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.

Metspalu et al. (2004) analyzed 796 Indians and 436 Iranians for mtDNA variation and compared with previous studies from India and also from Europe, China and Thailand. The northern states (caste populations) showed higher frequencies of western Eurasian haplogroups. The west Eurasian haplogroups can be broadly divided into two groups, one showing admixture within the last 10000 years and the other (one third) showing much deeper time depths (40000 years). Tibeto-Burman speaking tribal populations of eastern and northern India exhibits the highest frequencies of East Eurasians specific mtDNA haplogroups. They concluded that deep autochthonous history of these haplogroups in the region remains to be the most parsimonious explanation. They propose that the initial mtDNA pool established upon the peopling of South Asia has not been replaced but has rather been reshaped in situ by major events of gene flow both from the west and the east during more recent chapters of the demographic history of the region

Sahoo and Kashyap (2006) reported 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. The mitochondrial data show hierarchical association with the Indo-European speakers of Eastern Europe.

Sun et al. (2006) performed complete genome sequencing of 56 mtDNA which covered all the recognized M lineages. They found that M Macrohaplogroup is ubiquitous in South Asia and covers more than half of Indian mtDNA. Only previous attempt at complete sequencing of mtDNA for M haplogroup was that of Rajkumar et al. (2005). Sun et al. (2006) points to errors typical of large mtDNA sequencing in their data. They found that the basal variation in macrohaplogroup M in India clearly outnumbers that of macrohaplogroup M in East Asia. The mtDNA phylogeny of M haplogroup is in good agreement with the proposed scenario that the initial dispersal of modern humans into Eurasia ~ 60K years ago was rather rapid along the Asian Coast line. Using 11 whole mtDNA and 2231 partial coding sequence of Indian M lineage selected from 8670 HVS1 sequences across India, Thangaraj et al. (2006) defined one novel haplogroup M41 and revised the classification of haplogroups M3, M18 and M31. Other haplogroups were also further classified into subhaplogroups. Phylogenetic tree was constructed including the data of Sun et al. (2006). Most of the new M lineages were found to be deep rooted and more likely arose in situ in Indian-subcontinent. The autochthonous lineages emerge directly from the root of the macrohaplogroup M. The deep rooting lineages are not language specific and are spread over all the language groups in India.

Palanichamy et al. (2004) did complete sequencing of 75 individuals belonging to macrohaplogroup N, to study the phylogenetic relationship of Indian and Western Eurasian mtDNA. They found that Indian N macrohaplogroup consists of some lineages which are as deep rooted as the western Eurasian mtDNA lineages. Also, 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. They conclude that these results along with the evidence for M macrohaplogroup suggest three-founder-mtDNA lineages migrating into the subcontinent at different time periods.

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 1 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 hitherto 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).

There have been a few studies focusing on the affinities of Indian Muslim populations the largest religious minority of India, as well. Terreros et al. (2007) studied 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.

Y-CHROMOSOME VARIATION IN INDIA

The paternally inherited Y-Chromosome provides insights into the origin, history and migratory pathways of the male lineages and helps reconstruct 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 AP (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 hypergyny played a role in it. Subsequently, based on a few STR markers, Bhattacharya et al. (1999) found that different ethnic groups of India represented in this study primarily by certain Indo-European and Austroasiatic linguistic groups harbor disjoint sets of haplotypes. Also, 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 dispersed 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 that 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 correlated 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 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 studied. 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).

Studies focusing on relative affinity of Indian caste and tribal gene pools with that of Asians and Europeans have also reported contrasting results for mtDNA and Y-Chromosomes. Mukherjee et al. (2001) in a study of 18 Y-Chromosomal polymorphic markers in 4 ethnic populations of Northern India reported higher frequencies of haplogroups HG-3 and HG-9, which are known to have arisen in central Asian region. NJ tree based on Y-Chromosome frequencies showed that North Indians are genetically placed between west Asian and central Asian populations.

In Contrast to mtDNA variation, Y-Chromosome variation in each of eight castes from Andhra was found to be more similar to East Europeans than to Asians (Bamshad et al. 2001). The affinity of Indian castes to Europeans was observed to be proportionate to caste rank. Based on mtDNA variation Bamshad et al. (2001) concluded that Indian castes are likely to be of proto-Asian in origin and the Y-Chromosome variation indicates west Eurasian admixture which is proportionate to the caste rank.

Some studies reported similar results for both mtDNA and Y-Chromosome variation. Kivisild et al. (2003) found that the major Indian Y-Chromosome haplogroups H, L and R2 occur in both the castes and tribal populations and 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, thus the results of both mtDNA and Y-Chromosome variation suggest a common genetic heritage of Indian castes and tribes.

Thanseem et al. (2006) analyzed three tribal populations from Andhra Pradesh and compared results with other populations. They found that in contrast to mtDNA, Y-Chromosome variation in India is distinct among caste and tribal populations. The lower castes showed 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 compared to upper castes prompting them to infer tribal origin of the lower castes.

Sahoo et al. (2006) typed 38 single-nucleotide polymorphic markers in 936 Y chromosomes, representing 32 tribal and 45 caste groups from all four major linguistic groups of India. Variation in the major Y-chromosomal haplogroups in India 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) analyzed 69 Y-chromosome SNPs and 10 microsatellite markers from a large set of geographically, socially, linguistically and ethnically diverse groups of South Asia and 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. Their results also support the deep antiquity of Indian Populations.

Basu et al. (2003) reported that Austro-Asiatic tribal groups possess high frequencies of Y-chromosomal haplogroup K* which is found in higher frequencies in Chinese and Southeast Asians populations. They concluded that Austro-Asiatic tribal populations entered India first from the Northwest corridor and much later some of them through Northeastern corridor. On the other hand, Kumar et al. (2007) analyzed 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 (Fig. 1). The conclusions 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 a barrier. Nevertheless, Kumar et al. (2007) and Basu et al. (2003) differ in the direction of migration for Austro-Asiatic populations through the Northeast corridor.

Cordoux et al. (2004b) analyzed Y-Chromosome data from 155 individuals from 9 tribal groups

and one caste population and compared with published data. The total dataset consisted of 931 individuals from 15 tribal and 12 caste groups. They found that seven most frequent haplogroups account for 80-90% of both caste and tribal samples, suggesting extensive overlapping of caste and tribal chromosomes. However, the castes and tribes differed in the frequency of the major haplogroups. Haplogroup O-M95 showed higher frequency in tribal than in caste groups. Caste groups were homogenous and were more closely related to Central Asian groups than to Indian tribal groups. From these results they concluded that caste and tribal populations of India have Independent origin. The sharing of most haplogroups was explained by admixture. They provide a very recent estimate for the time (3500 years) for migration of Indo-European speakers from central Asia from whom the Indian caste populations are supposed to have been derived.

Kumar et al. (2006b) tested the hypothesis that Y-Chromosome variants tend to be more localized geographically than those of mtDNA variants because of the widespread phenomenon of patrilocality (Seielstad et al. 1998). This hypothesis is based on the model of isolation by distance and it got some support from studies in Thailand (Oota et al. 2001). Kumar et al. (2006b) tested the universality of this hypothesis by analyzing Y-chromosome and mtDNA data in three different sets of Indian populations that follow endogamy rules to varying degrees. The results showed that the Indian patrilocal and the matrilocal groups do not conform to the sex-specific variation as observed among the tribes of Thailand. The patterns of genetic variability in India are not consistent with the above hypothesis as the population structure is unique based on the endogamy rules, hence adhere to the island model.

Redd et al. (2002) showed additional 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 analyses of STR variation associated with a major Australian SNP lineage indicated tight clustering with southern Indian/Sri Lankan Y chromosomes. Estimates of the divergence for 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.

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 East Asia, Central Asia, India, Sri Lanka, Pakistan, Iran, the Middle East, Turkey, Egypt and Morocco. They found that the Muslim populations in general are genetically closer to their non-Muslim neighbors than to other Muslims in India. They conclude that the Muslim expansion in India was predominantly a cultural change and was not accompanied by significant gene flow. The conclusions are similar to what was observed for mtDNA among the Muslims of Uttar Pradesh (Terreros et al. 2007). Basu et al. (2003) have also found the Muslims of Uttar Pradesh to cluster with Indo-European Upper-Caste groups for Y-Chromosome and autosomal markers.

AUTOSOMAL VARIATION IN INDIA

As against mtDNA and Y-Chromosome which show a simple pattern of inheritance, the autosomal DNA has not been a favorite 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.

Majumder et al. (1999) analyzed DNA samples from 396 individuals belonging to 14 ethnic populations of India for 8 human specific polymorphic insertion/deletion loci and observed that geographically closer populations showed greater affinity than populations with socio-cultural similarity. In contrast, Roychoudhury et al. (2001) observed genomic affinity to show good association with linguistic affinity for autosomal loci. Viswanathan et al. (2004), on the other hand, found that genetic distances based on 24 autosomal markers were not correlated with geographic distances for 5 Dravidian speaking tribal populations from Nilgiri hills area of Tamil Nadu. They also found that the Indian populations were closely related to each other, regardless of the fact

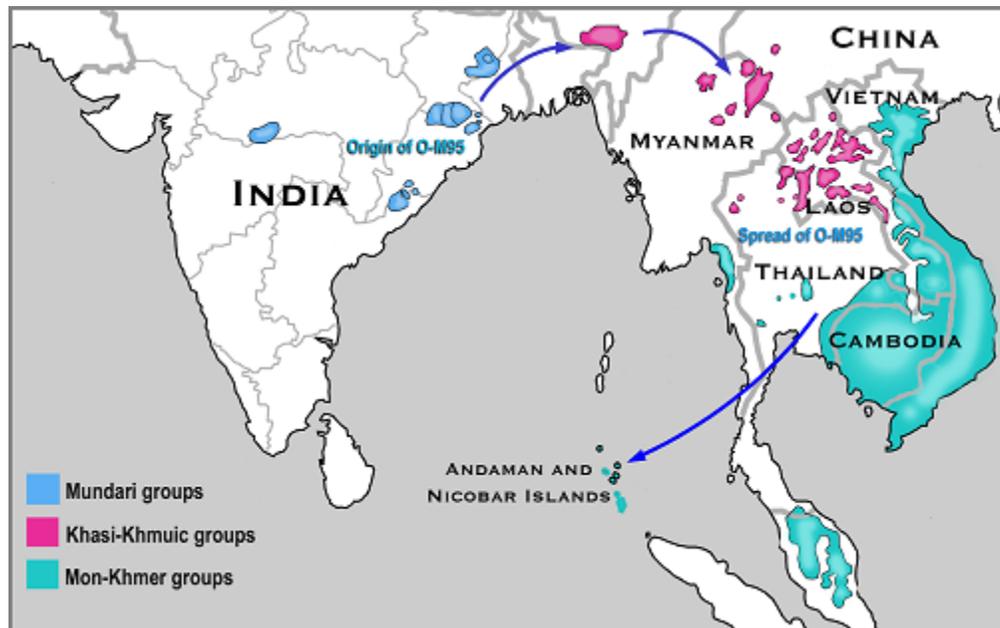


Fig. 1. Map showing the distribution of Austro-Asiatic populations and the origin of haplogroup O-M95 among the Mundari and the route of its migration.

(Source: Kumar et al. 2007)

that some tribal populations showed Negrito phenotypic characteristics. Basu et al. (2003) observed that the clustering (based on mtDNA, Y-Chromosome and autosomal markers) of populations was not strictly consistent with their social, geographical or linguistic affiliations, except that the Tibeto-Burman speakers of Northeast form a separate cluster. Population structure analysis also showed that Indo-European speakers and Dravidian speakers are the most similar for Autosomal data, though these two groups are geographically wide apart.

Autosomal STRs and Population Structure at Micro and Macrolevels

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, Kritika 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 (Fig. 2). 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) analyzed 9

populations, suggesting genetic identity of these populations. Overall, the observation of Reddy et al. (2005) 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 Chittoor district of AP. 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 heterogeneous geographic background were studied.

Kivisild et al. (2003) found that frequency of autosomal haplogroups MX1 and Ch21 distinguished the Koyas and Chenchus, two tribes from South India, along with other Indian caste populations, from the European and Eastern Asian populations. This result corroborates with their mtDNA and Y-Chromosome based conclusion suggesting 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. 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 homogeneity although still the contemporary caste and tribal groups formed distinct clusters in both Principal Component plot and Neighbor 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 to 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 population 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 neighboring 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).

Studies Focusing on Populations of Andaman and Nicobar Islands

The issue of the origin of Andaman Nicobar tribes has always been intriguing to Anthropologists. These tribes have been proposed to provide foot prints of migration of early wave of modern humans to Australia (Redd et al. 2002; Macaulay 2005). A few recent studies have focused on the molecular genetic features of populations from Andaman Nicobar islands. Analyzing mtDNA sequences and RFLP polymorphisms, Y-Chromosome biallelic marker and microsatellites in the populations of Andaman and Nicobar islands, Thangraj 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 analyzed 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- 70 kya. 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, Thangraj 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) analyzed mtDNA hypervariable region 1 sequence data from 33 Nicobarese Islanders and compared their mtDNA haplotypes to those of neighboring East Asians, mainland and island Southeast 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) analyzed 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 analyses 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 analyses 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 descendents 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 at 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 Courdaux and Stoneking (2003) differ with this conclusion and opine that the southern route hypothesis should await adequate genetic support.

EMERGING PICTURE

Most of the studies based on mtDNA variation have reported genetic unity of Indian populations. The basic clustering of maternal lineages has been reported to be not specific to a particular language or caste (Mountain et al. 1995; Kivisild et al. 1999; Bamshad et al. 2001). More than 60% of the Indians have their maternal roots in Indian specific branches of haplogroup M. Because of its great time depth and virtual absence in western Eurasians, it has been suggested that haplogroup M was brought to Asia from East Africa along the southern route by earliest migration wave of AMH ~ 60000 years ago (Mountain et al. 1995; Kivisild et al. 1999, 2000; Quintana Murci et al. 1999; Sun et al. 2006). Haplotype U was found to connect western Eurasian and Indian populations. Both macrohaplogroup M and N suggest deep ancestry of Indian population dating back to 40,000-60,000 years (Metspalu et al. 2004 and Palanichamy et al. 2004). More than one wave of migration of people

bearing West Eurasian haplogroups from west Asia into India has been reported.

Cordaux et al. (2003) reported that mtDNA variation indicated strong affinity of Indian tribal and caste groups to east Eurasians, while Bamshad et al. (2001) and Barnabas et al. (1996) reported that west Eurasian haplotypes were found to represent about a quarter of Caste mtDNA haplotypes. Baig et al. (2004), Kivisild et al. (2003), Cordaux et al. (2003), Roychoudhury et al. (2000), reported that both caste and tribal groups share the same maternal lineages. In contrast, Watkins et al. (1999) based on mtDNA 9bp deletion reported different origins for caste and tribal populations, whereas Kumar et al. (2006) inferred Asian and non-Asian origins of MonKhmer and Mundari groups of Austro-Asiatic tribes. Some of the studies have reported correlation of social rank with the mtDNA distances among the caste populations (Thanseem et al. 2006; Sahoo and Kashyap 2006). Relatively greater affinity of north-west Indian populations to West Eurasians has also been observed (Passarino et al. 1996a; Bamshad et al. 1996, 2001; Quintana-Murci 2004). High mtDNA diversity and deep coalescence dates (~56-63k ybp) suggest that Austo-Asiatic speaking groups might be original inhabitants of

India (Kumar et al. 2006; Basu et al. 2003; Roychoudhuri et al. 2001; Kivisild et al. 1999).

Overall, thus, the studies based on mtDNA suggest deep rooted maternal lineages for both caste and tribal populations 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.

In contrast to mtDNA variation the conclusions based on Y-Chromosome variation have been quite varied. Ramana et al. (2001), Thanseem et al. (2003), Cordaux et al. (2004b) reported that the patterns of variation based on frequencies of some haplogroups are distinct among castes and tribal groups. While Bamshad et al. (1998) and Bhattacharya et al. (1999) reported very little or no paternal gene flow among Indian populations, Ramana et al. (2001) suggest possibility of male gene flow between tribal and caste populations. Concurrent to this, Kumar et al. (2006) observed that Y-Chromosome variants are not localized

Table 2: Coalescence age for different haplotypes and populations in different studies in Indian populations

| Diagnostic Coding Region Marker | Coalescence (years) | Coalescence years (X1000) | Population | Expansion times/ TMRCA (confidence interval) |
|---------------------------------|------------------------------|---------------------------|-----------------|--|
| M2 | 70,600 ± 21,000 ¹ | | Austro-Asiatic | 56098 (51220-60975) ² |
| M2a | 48,300 ± 20,100 ¹ | 70.2 ³ | Dravidian | 39024 (34146-43902) ² |
| M2b | | 77.1 ³ | Tibeto-Burman | 51220 (48780-53659) ² |
| M3 | | 24.0 ³ | | |
| M3a | 17,300 ± 7,400 ¹ | | | TMRC A |
| M4a | 19,200 ± 9,000 ¹ | | Munda | 65,730 (25,442-132,230) ⁵ |
| M5* | | 46.3 ³ | Khasi | 57,252 (27,644-92,201) ⁵ |
| M6 | 33,000 ± 13,900 ¹ | | Nicobarese | 16,578 (4,565-51,377) ⁵ |
| M6a | 19,100 ± 7,600 ¹ | | Austro-Asiatic | 68,098 (25,992-146,833) ⁵ |
| M6b | 6,000 ± 2,100 ¹ | 92.5 ³ | | |
| M18 | 9,400 ± 3,200 ¹ | 41.1 ³ | | Age estimate of M Haplogroup |
| M25 | 19,400 ± 7,200 ¹ | | | |
| M33 | | 43.7 ³ | South Asia | 44600 ± 3300 ³ |
| M35a | | 36.0 ³ | East Asia | 69300 ± 5400 ³ |
| M38 | | 54.0 ³ | Oceania | 73000 ± 7900 ³ |
| M39 | | 39.1 ³ | South East Asia | 55700 ± 7400 ³ |
| M40 | | 38.6 ³ | | |
| M* | | 41.1 ³ | | Coalescence age for Maharastrian populations |
| Total M | | 54.1 ³ | | |
| R2e | 40,400 ± 14,300 ¹ | | M Total | 45082 ± 641 ⁴ |
| R5 | 66,100 ± 22,000 ¹ | | U Total | 25628 ± 1624 ⁴ |
| R6 | 30,000 ± 11,000 ¹ | | R Total | 45001 ± 1039 ⁴ |

¹ Metspalu 2004, ² Majumder 2001, ³ Sun et al. 2006, ⁴ Baig et al. 2004, ⁵ Kumar et al. 2007

geographically and also suggest the possibility of male gene flow. Kivisild et al. (2003), on the other hand, suggested ancient and shared genetic heritage of male lineages in India. Genetic affinity with west Eurasian gene pool was found to be correlated with caste rank by Bamshad et al. (1998) and Sahoo and Kashyap (2006). Bamshad et al. (2001) further suggested that this affinity with west Eurasians is proportionate to caste rank. While Mukherjee et al. (2001) placed North Indians between west Asian and Central Asian populations, Cordaux et al. (2004b) placed Indian caste populations closer to Central Asian populations. On the other hand, while Bamshad et al. (2001) placed caste populations of southern Indian states closer to East Europeans, Sahoo et al. (2006) and Sengupata et al. (2006) suggest no recent admixture in the Indian caste populations. This conclusion is similar to most of the mtDNA studies. However, Cordaux et al. (2004b) infer a relatively recent migration of Indo-European speakers to India lending support to the Aryan invasion/migration theory.

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 distances 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. 2004; 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. 2005; Majumder et al. 1999) the clusters were formed consistent to geographic affiliation of the groups. Roychoudury 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 of 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.

With the emergence of molecular genetic techniques the general expectation had been that the riddles of physical anthropology will be solved. Indeed, these techniques have been extensively used to study human evolution and variation and to trace the routes of migration of the humans from Out of Africa to the rest of the world, in conjunction with the data from paleo-anthropology and archaeology. One of the advantages of these studies has been that the Time of Most Recent Common Ancestor (TMRCA) could be estimated using more robust assumptions in comparison to the traditional markers. However, these studies have their share of pitfalls as well; for example, the TMRCA estimated from quite a few studies have been found to be less than the time estimated from the fossil or archeological finds. Nevertheless, at the global level, many of the finer details of major population relationships and routes of migration have been elucidated with fair degree of confidence.

At micro-level, molecular anthropological studies on Indian populations have indeed come up with quite interesting but sometimes with contrasting results. Some of the results like common and deep genetic heritage of Indian caste and tribal populations have strongly contradicted the previous hypothesis of recent entry of Indo-European speakers in India. Many studies have also come up with dates of common ancestry and/or for initial colonization of Indian subcontinent as ~60,000 years which roughly corresponds with the first wave of migration of humans out of Africa. Some of the studies substantiated the findings of classical markers, like the greater affinity of caste populations of higher social hierarchy to the Europeans and also that of geographically closer population are genetically closer as well.

In spite of these interesting results, there have been quite contrasting conclusions from different studies. Contrasting conclusions on population affinities and coalescence dates from the common ancestors are reported in different studies. Table 2 presents the age of coalescence for different haplogroups and populations from different studies. Although, the dates seem quite interes-

ting, they are accompanied by very large confidence intervals which make the conclusions equivocal. Also the same haplotype is assigned quite contrasting dates in different studies. A critical appraisal of the literature makes one to wonder what is going wrong. Are there some grave methodological problems or is it because the Indian population structure is too complex to be described in simple terms as different studies have attempted to do? One striking aspect of these molecular anthropology studies is the fact that very often the sweeping conclusions are made with inappropriate study frame, compounded by inadequate samples and sampling. Often the sampling details and/or population details were not furnished and one is prompted to surmise if the investigators really knew and/or care about these aspects which are crucial in population genetics research. Sometimes, the samples used for the study does not even match the requirement of the hypothesis being tested. For example, including geographically and ethnically/linguistically disparate populations without even remote possibility of exchange of genes, either currently or historically, to test the hypothesis of male gene flow between them is far from being a suitable study frame given that India offers immense possibilities to find appropriate situations to test many such hypotheses. Identification of suitable population units has also been a problem in some studies; populations groups as diverse as Hindi speaking, North Indian, etc have been used as units of study. Quite a few studies lack proper description of the populations investigated, which is of great importance in such studies. One review in fact wrongly defines Scheduled caste (SC) population and then describes the tribal population 'as most disadvantaged group in SC' (Chaubey et al. 2006). One of the major reasons behind these 'mishaps' might be lack of anthropological insights into Indian population structure, as many of the papers have been written by people of non-Anthropology (especially Indian Anthropology) background. This may be the reason why we find one of the studies describing the process of Hypergamy as a process which has resulted in similar maternal gene pools of the Indian population. The process of Hypergamy though described in religious text and other records has not been reported to be a commonly practiced by the masses to cause such systematic genetic patterns.

The importance of population based approach has been realized in the emerging fields of

Genetic Epidemiology and pharmacogenetics where the focus is on identifying genes associated with complex diseases and in turn use this information in individual and/or population specific treatment. Thus, a proper knowledge of population and/or population structure is crucial in all forms of molecular genetic research. Population based approach has been the hallmark of anthropological research from the beginning. Anthropologists can thus make a major contribution in the emerging research fields of molecular genetics. Unfortunately anthropologists have been marginalized because of the advent of sophisticated and resource intensive technologies of modern genetics. It is highly imperative that fruitful research can be conducted if molecular biologists practicing anthropological genetics work in mutual collaboration with biological anthropologists.

ACKNOWLEDGEMENT

The authors are thankful to the Director, Indian Statistical Institute, Kolkata and Chairman, Mahatma Gandhi National Institute of Research and Social Action (MGNIRSA), Hyderabad for logistic support.

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. mtDNA variation in caste populations of Andhra Pradesh, India. *Hum Biol*, **68**: 1-28.
- Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BVR, Reddy PG, Rasanayagam A, Papiha SS, Villems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, Jorde LB 2001. Genetic Evidence on the Origins of Indian Caste Populations *Genome Res*, **11**: 994-1004.
- 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.
- 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.
- Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S,

- Chakraborty M, Dey B, Roy M, Roy B, Bhattacharyya NP, Roychoudhury S, Majumder PP 2003. Ethnic India: A Genomic View, With Special Reference to Peopling and Structure. *Genome Res*, **13**: 2277-2290.
- Bhasin MK, Walter H, Danker-Hopfe H 1992. *The Distribution of Genetical, Morphological and Behavioural Traits Among the Peoples of Indian Region*. Delhi: Kamla-Raj Enterprises.
- Bhasin MK, Walter H, Danker-Hopfe H 1994. *People of India. An Investigation of Biological Variability in Ecological, Ethno economic and Linguistic Groups*. Delhi: Kamla-Raj Enterprises.
- Bhasin MK, Walter H 2001. *Genetics of Castes and Tribes of India*. Delhi: Kamla-Raj Enterprises.
- 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 Polymorphisms. *Genome Res*, **9**: 711-719.
- Buxton LHD 1925. *The Peoples of Asia*. London: Kegan Paul, Trench & Trübner.
- Cann RL, Stoneking M, Wilson AC 1987. Mitochondrial DNA and human evolution.
- Cavalli-Sforza LL, Menozzi P, Piazza A 1994. *The History and Geography of Human Genes*. Princeton: Princeton University Press.
- Chaubey G, Metspalu M, Kivisild T, Villems R 2006. Peopling of South Asia: investigating the caste-tribe continuum in India. *Bio Essays*, **9**: 91-100.
- 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 populations of India. *Hum Biol*, **72**: 273-285.
- 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.
- Cordaux R, Saha N, Bentley GR, Aunger 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, 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, Weiss G, Saha N, Stoneking M 2004a. The northeast Indian passageway: a barrier or corridor for human migrations? *Mol Biol Evol*, **21**: 1525-1533.
- 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.
- 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.
- Gadgil M, Malhotra KC 1983. Adaptive significance of the Indian caste system: An ecological perspective. *Annals of Human Biology*, **10**: 465-478. pp 295-324.
- 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.
- Gutala R, Carvalho-Silva DR, Jin L, Yngvadottir B, Avadhanula V, Nanne K, Singh L, Chakraborty R, Tyler-Smith C 2006. A shared Y-chromosomal heritage between Muslims and Hindus in India. *Hum Genet*, **120**: 543-551
- 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, 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.
- 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.
- 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, Gölge M, Usanga E, Papiha SS, Cinnioğlu C, King R, Cavalli-Sforza L, Underhill PA, Villems R 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet*, **72**: 313-332.
- 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 V, Langstieh BT, Biswas S, Babu JP, Rao TN, Thangaraj K, Reddy AG, Singh L, Reddy BM 2006a. 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, Langstieh BT, Madhavi KV, Naidu VM, Singh HP, Biswas S, Thangaraj 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, Reddy AN, Babu JP, Rao TN, Langstieh BT, Thangaraj 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.
- Kumar V, Reddy BM 2003. Status of Austro-Asiatic groups in the peopling of India: An exploratory study based on the available prehistoric, linguistic and biological evidences. *J Biosci*, **28**: 507-522.
- 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.
- 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 2001. Ethnic populations of India as seen from an evolutionary perspective. *J Biosci*, **26**: 533-545.
- 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, Mukherjee BN 1993. Genetic diversity and affinities among Indian populations: An overview. In: PP Majumder (Ed.) *Human Population Genetics*. New York: Plenum pp. 255-275.
- Malhotra KC 1978. Morphological composition of the people of India. *J Hum Evol*, **7**: 45-53.
- Malhotra KC 1984. Population structure among the Dhangar Caste-cluster of Maharashtra, India. In: JR Lukacs (Ed.) *The People of South Asia*. New York: Plenum Press.
- Malhotra KC, Vasulu TS 1993. Structure of Human populations of India. In: PP Majumder (ed.) *Human Population Genetics*. New York: Plenum Press. pp. 207-275.
- Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Vilems R 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet*, **5**: 26.
- Mitchell RJ, Reddy BM, Campo D, Infantino T, Kaps M, Crawford MH 2006. Genetic diversity within a caste population of India as measured by Y-chromosome haplogroups and haplotypes: subcastes of the Golla of Andhra Pradesh. *Am J Phys Anthropol*, **130**: 385-393.
- Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadjil M, Cavalli-Sforza LL 1995. Demographic history of India and mtDNA-sequence diversity. *Am J Hum Genet*, **56**: 979-992.
- Mukherjee N, Nebel A, Oppenheim A, Majumder 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.
- Oota H, Settheetham-Ishida W, Tiwaweeh D, Ishida T, Stoneking M 2001. Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nature Genet*, **29**: 20-21.
- Palanichamy MG, Sun C, Agrawal 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-Benerecetti AS 1996a. Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. *Am J Hum Genet*, **59**: 927-934.
- Passarino G, Semino O, Modiano G, Bernini LF, Santachiara Benerecetti AS 1996b. mtDNA provides the first known marker distinguishing proto-Indians from the other Caucasoids; it probably predates the diversification between Indians and Orientals. *Ann Hum Biol*, **23**:121-126.
- Pattanayak DP 1998. The language heritage of India. In: NA Rao (Ed.) *The Indian human heritage*. Hyderabad: Universities Press pp. 95-99.
- 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-Benerecetti 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, Banerjee J, Gunturi HB, Trivedi R, Kashyap VK 2005. Phylogeny and antiquity of M macrohaplogroup inferred from complete mtDNA sequence of Indian specific lineages. *BMC Evol Biol*, **5**: e26.
- 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.
- Ranjan D, Trivedi R, Vasulu TS, Kashyap VK 2003. Geographic contiguity and genetic affinity among five ethnic populations of Manipur, India: Further molecular studies based on VNTR and STR loci. *Ann Hum Biol*, **30**: 117-131.
- Rapson EJ 1955. Peoples and languages. In: EJ Rapson (Ed.) *Cambridge History of India, Vol 1: Ancient India*. Delhi: S Chand and Co.
- 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, Dutta R, Langstieh 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, Naidu VM, Madhavi VK, Thangaraj LK, Kumar V, Langstieh BT, Venkatramana P, Reddy AG, Singh L 2005. Microsatellite diversity in Andhra Pradesh, India: genetic stratification versus social stratification. *Hum Biol*, **77**: 803-823.
- Reddy BM, Langstieh BT, Kumar V, Nagaraja T, Aruna M, Thangaraj K, Reddy ANS, Reddy AG, Singh L 2007. Austro-Asiatic tribes of northeast India provide missing genetic link between south and southeast Asia, *PLoS ONE* (in press)
- Reddy BM, Sun G, Luis JR, Crawford MH, Hemam NS, Dekka 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.
- Reid RM 1984. Taxonomic distance and Human population of South Asia. In: JR Lukacs (Ed.) *The People of South Asia*. New York: Plenum Press. pp 269-281.
- Risley HH 1915. *The People of India*. Calcutta: Thacker Spink.
- Roychoudhury AK 1984. Genetic relationship of Indian population. In: KC Malhotra, A Basu (Eds.) *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Human Genetics and Adaptation, Vol. 2 Human Genetics*. Calcutta: Indian Statistical Institute.
- Roychoudhury S, Roy S, Basu A, Banerjee R, Vishwanathan 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.
- Roychoudhury S, Roy S, Dey B, Chakraborty M, Roy M, Roy B, Ramesh A, Prabhakaran N, Rani MVU, Vishwanathan 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.
- 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.
- Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, Gaikwad S, Trivedi R, Endicott P, Kivisild T, Metspalu M, Villems R, Kashyap VK 2006. A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc Natl Acad Sci USA*, **103**: 843-848.
- Sarkar SS 1958. Race and race movements in India. In: SK Chatterjee (ed.) *The cultural heritage of India*. Calcutta: The Ramkrishna Mission Institute of Culture pp. 17-32.
- Seielstad MT, Minch E, Cavalli-Sforza LL 1998. Genetic evidence for a higher female migration rate in humans. *Nature Genet*, **20**: 278-280.
- Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, Lin AA, Mitra M, Sil SK, Ramesh A, Usha Rani MV, Thakur CM, Cavalli-Sforza LL, Majumder PP, Underhill PA 2006. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central asian pastoralists. *Am J Hum Genet*, **78**: 202-221.
- Singh KS 1993. People of India: The profile of national project (1985-92). *Curr Sci*, **64**: 5-10.
- 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, Agrawal 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.
- Terreros MC, Rowold D, Luis JR, Khan F, Agrawal S, Herrera RJ. 2007. North Indian Muslims: Enclaves of foreign DNA or Hindu converts? *Am J Phys Anthropol*, **133**: 1004-1012.
- Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L 2005. Reconstructing the origin of Andaman Islanders. *Science*, **308**: 996.
- Thangaraj K, Chaubey G, Singh VK, Vanniarajan A, Thanseem I, Reddy AG, Singh L 2006. In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup M in India. *BMC Genomics*, **7**: 151.
- Thangaraj K, Sridhar V, Kivisild T, Reddy AG, Chaubey G, Singh VK, Kaur S, Agarawal P, Rai A, Gupta J, Mallick CB, Kumar N, Velavan TP, Suganthan R, Udaykumar D, Kumar R, Mishra R, Khan A, Annapurna C, Singh L. 2005. Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. *Hum Genet*, **116**: 507-517.
- Thangaraj K, Chaubey G, Reddy AG, Singh VK, Singh L 2006. Unique origin of Andaman Islanders: insight from autosomal loci. *J Hum Genet*, **51**: 800-804.
- Thangaraj K, Ramana GV, Singh L 1999. Y-chromosome and mitochondrial DNA polymorphisms in Indian populations. *Electrophoresis*, **20**: 1743-1747.
- 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.
- 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 Genomics*, **7**: 42.
- Thapar R 1966. *A History of India. Vol 1*. Middlesex: Penguin.
- Trivedi R, Sitalaximi 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.
- Tripathy V, Reddy BM 2007. Present Status of understanding on the G6PD and Natural Selection. *Journal of Postgraduate Medicine*, **53**: 193-202.
- Vishwanathan 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, Ramachandran 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.