

## Distribution of Mitochondrial DNA Macrohaplogroup N in India with Special Reference to Haplogroup R and its Sub-Haplogroup U

Suvendu Maji, S. Krithika and T.S. Vasulu\*

*Biological Anthropology Unit, Indian Statistical Institute, Kolkata 700 108, West Bengal, India*

**KEYWORDS** Mitochondrial DNA diversity; macrohaplogroup N; haplogroup R; sub-haplogroup U; Indian populations

**ABSTRACT** Indian populations show mitochondrial DNA macrohaplogroups M and N. Among the various haplogroups of N, haplogroup R and its sub-haplogroup U are predominant among the diverse populations. An overview of about 130 studied populations reveals wide diversity of haplogroup R and its sub-haplogroup U among varied ethnic status and different linguistic families. The various lineages of macrohaplogroup N observed in India shows least affiliation with respect to language, geography or ethnicity. However, between the regions, there is a wide variation of subhaplogroups of R and U; western region among the castes and southern region among the tribes show higher haplogroup diversity than the other regions.

### INTRODUCTION

Mitochondrial DNA (mtDNA), owing to its unique properties such as high copy number, near-absence of genetic recombination, maternal inheritance and high mutation rate (3 to 5 times greater than nuclear DNA), has proven to be a powerful and an efficient molecular genetic tool for investigating the (maternal) genetic history of human populations (Cann et al. 1987). In the beginning, mtDNA studies involved the restriction site polymorphisms (RFLPs) and the non-coding/control regions sequences (hypervariable region I and II) (Brown 1980; Danero et al. 1981; Bamshad et al. 2001; Basu et al., 2003; Kivisild et al. 1999) for categorizing the individuals into specific haplogroups (related groups of sequence defined by shared mutations) based on the mutations observed in relation to the reference sequence (Anderson et al. 1981; Andrews et al. 1999). However, these polymorphisms neither provide a detailed characterization of the haplogroups nor complete phylogenetic information of the populations. Later, high-throughput sequencing technologies were employed for complete or near complete mitochondrial genome sequencing for characterization of the phylogeny of mtDNA haplogroups to infer the possible origin and

migration of man out of Africa and their settlements in different continents (Ingman et al. 2000; Finila et al. 2001; Maca-Mayer et al. 2001; Kong et al. 2003; Palanichamy et al. 2004).

The phylogenetic studies of mtDNA diversity of world populations indicate that L is the oldest macrohaplogroup, restricted to African populations, and consists of seven major branches or haplogroups: L0, L1, L2, L3, L4, L5, and L6 (Cann et al. 1987; Watson et al. 1997; Cann 2001; Salas et al. 2002; Kivisild et al. 2004; Kivisild et al. 2006). Of the major African mtDNA haplogroups, one branch (haplogroup L3) emigrated out of Africa and dispersed ~ 60,000 ybp into Eurasia and the New World, in the form of superclades or macrohaplogroups M and N, (Quintana-Murci et al. 1999; Mishmar et al. 2003). M was initially defined (Chen et al. 1995) by the presence of two closely located restriction enzyme sites, *DdeI*<sup>10394</sup> (10398 A to G) and *AluI*<sup>10397</sup> (10400 C to T) and was first detected among South-East Asian populations (Ballinger et al. 1992). Macrohaplogroup M encompass the Asian-specific haplogroups C, D, E, G and Z. Macrohaplogroup N was defined by presence of *DdeI*<sup>10394</sup> site and absence of *AluI*<sup>10397</sup> site and includes A, B, F and Y haplogroups. N extends its branches all over Eurasia and segregate to the eastern (*e.g.* sub-haplogroups A, B, Y and R9) or western Eurasian-specific pools (*e.g.* sub-haplogroups N1, N2, TJ, HV, U).

The hypothesis of out-of-Africa-migration proposes that the earliest migrants, possibly

\*Address for correspondence: Dr. T. S. Vasulu, Biological Anthropology Unit, Indian Statistical Institute, 203 BT Road, Kolkata 700108, West Bengal, India.  
Telephone: 91 33 25753215  
E-mail: vasulu@gmail.com, vasulu@isical.ac.in

carried mtDNA macrohaplogroups M and N, entered southern Asia and later some splinter groups have drifted to Southeast Asia and reached Australia (Cann 2001; Macaulay et al. 2005) leaving their genetic signatures all along their route of migration. These findings also highlight the importance of Indian subcontinent in the early genetic history of human settlement and expansion. The past genetic history of Man can be understood by study of mtDNA and Y-chromosome polymorphisms among the extant autochthonous Indian tribes. Besides, the molecular genetic studies will also provide impetus to readdress some of the basic issues related to ancient population structure of early settlers who developed hierarchical social structure of different castes, its biological implications and patterns of diversity among castes and tribes etc.

In India various regional castes, tribes and religious groups have been studied for mtDNA diversity. These studies reveal interesting insights into the complex mtDNA phylogenetic tree and the genetic structure of Indian populations. The macrohaplogroups M and N are prevalent across different regions, among both castes and tribes, and also show some Indian specific subclades or subhaplogroups. In the present study, we describe the distribution of various haplogroups and subhaplogroups of macrohaplogroup N among about 130 populations studied in India (Fig. 1) and investigate the patterns of mtDNA diversity across regions, castes, tribes, and linguistic groups.

#### Distribution of N and Its Subhaplogroups

Overall, a collation of mtDNA diversity studies, especially for N and its subhaplogroups, indicates that about 70 castes and 60 tribes from six different regions have been studied. About 37 per cent (highest representation) of the studied populations are from southern region whereas the least (3.8%) are from Islands.

Nine subhaplogroups of N are reported among the studied populations of which haplogroup R shows about 15 branches and U the subhaplogroup shows 8 variants. In India, most of the sub-haplogroups of N are the region (Indian) specific variants of haplogroup R (characterized by motif: 12705, 16223) including R5, R6 and U2 (a, b, c) (Metspalu et al. 2004). The coalescence age of the Indian 'N' cluster is estimated to be  $45,405 \pm 7,752$  ybp (Barnabas et al. 2005). Region-wise distribution of macrohaplogroup N in the

Indian subcontinent is represented in Figure 2. In the Indian scenario, since haplogroup R and its sub-haplogroup U are predominant, we primarily focus on haplogroup R and the distribution of its sub-haplogroups (especially U) across the subcontinent. A schematic representation of the classification of haplogroup R and its sub-haplogroup U in India is depicted in Figure 3.

### HAPLOGROUP R

#### Sub-haplogroup U

Kivisild et al. (1999) studied 550 Indian mtDNA samples, chiefly obtained from Uttar Pradesh (Tharu, Buksa, Mixed castes), Punjab (Lobana) and Andhra Pradesh (Lambadi, Mixed castes), by sequencing hypervariable region I (HVI) and by performing RFLP analysis. Among the observed haplogroups, U was found to be the second most prevalent haplogroup (~ 13%) in India, being more common (~ 23%) in north India (Uttar Pradesh, n = 103) than in south India, ~ 10% (Andhra Pradesh, n = 250). The principal sub-haplogroup of U in India is U2 (named U2i), which lacks a characteristic transversion at np 16129 and hence differs from the Western-Eurasian variety of U2 (named U2e). In India, U2i exhibited a frequency of about 78% of the total, U followed by U7 that showed a frequency of about 13%. The split between Indian and Western-Eurasian U2 lineages was estimated to be  $53,000 \pm 4,000$  ybp and the coalescence age of U7 was calculated to be  $32,000 \pm 5,500$  years. Other subclusters of U found at lower frequencies were U1 (2.3%), U4 (4.7%), U5 (1.2%) and K (1.2%). Roychoudhury et al. (2000) reported an overall frequency of 14.28% for haplogroup U in his study including 644 mtDNA samples from 23 ethnic populations of India. Samples, collected from eastern (West Bengal, Orissa and Tripura), central (Madhya Pradesh), northern (Uttar Pradesh) and southern (Tamil Nadu) regions of the subcontinent, were screened for seven RFLP markers (*HaeIII* np 663, *AluI* np 5176, *DdeI* np 10394, *AluI* np 10397, *HinfI* np 12308, *HincII* np 13259 and *HaeIII* np 16517) and one insertion/deletion polymorphism (COII/tRNA<sup>Lys</sup> intergenic 9-bp deletion). Populations of northern India (Muslims (32.2%), Chamar (32%) and Brahmins (25.9%) of Uttar Pradesh) were found to exhibit higher frequencies of haplogroup U than populations of southern India. Also, Austro-Asiatic



Map not to Scale

Fig. 1. Map of India showing the location of the populations studied for mitochondrial DNA polymorphisms with special reference to macrohaplogroup N

1- Mixed Caste, 2- Kanet, 3- Jat Sikh, 4- Scheduled Caste, 5- Punjab Brahmins, 6- Lobanas, 7- Mixed Caste, 8- Rohidas, 9- UP Brahmins, 10- Chamar, 11- Muslims, 12- Rajput, 13- Khatri, 14- Tharu, 15- Buksa, 16- Bhargava, 17- Chatruvedi, 18- Mixed Caste, 19- Yadava, 20- Tiwari, 21- Adi, 22- Apatani, 23- Nishi, 24- Naga, 25- Meitei, 26- Mizo, 27- Tipperah, 28- Chakma, 29- Jamatia, 30- Mog, 31- Riang, 32- Khasi, 33- Garo, 34- Lodha, 35- Santal, 36- Munda, 37- Bagdi, 38- WB Brahmins, 39- Mahiashya, 40- Toto, 41- Rajbhansi, 42- Mixed Caste, 43- Kurmi, 44- Agharia, 45- Gaud, 46- Juang, 47- Munda, 48- Tanti, 49- Saora, 50- Paroja, 51- Gope, 52- Karan, 53- Bhuiyan, 54- Bathudi, 55- Khandayat, 56- Orissa Brahmins, 57- Gadaba, 58- Ho, 59- Gujrati, 60- Mixed Caste, 61- Kathodi, 62- Desasth Brahmins, 63- Chitpavan Brahmins, 64- Dhangar, 65- Maratha, 66- Konkanastha Brahmins, 67- Naba Baudh, 68- Other Maharashtrian Brahmins, 69- Parsi, 70- Rajput, 71- Gond, 72- Koli, 73- Muria, 74- Kanwar, 75- Kamar, 76- Halba, 77- Saryupari Brahmins, 78- Kadar, 79- Toda, 80- Kuruchian, 81- Mullukurunan, 82- Paniyan, 83- Mixed Caste, 84- Cochin Jews, 85- Havik, 86- Mukri, 87- Soliga, 88- Koraga, 89- Yerava, 90- Karnataka Brahmins, 91- Irula, 92- Kota, 93- Kurumba, 94- Ambalakarar, 95- Iyer, 96- Iyengar, 97- Pallan, 98- Vanniyar, 99- Kadar, 100- Paniyan, 101- Vellala, 102- Jennukurumba, 103- Kattunaikan, 104- Mullukurumba, 105- Bettakurumba, 106- Baduga, 107- Niyogi Brahmins, 108- Vydiki Brahmins, 109- Kshtriya, 110- Vyaya, 111- Teliga, 112- Turpu Kapu, 113- Yadava, 114- Relli, 115- Madiga, 116- Mala, 117- Pardhan, 118- Naikpod, 119- Chenchu, 120- Koya, 121- Lambadi, 122- Pardhi, 123- Thoti, 124- AP Brahmins, 125- Reddy, 126- Thogataveera, 127- Jarwa, 128- Onge, 129- Great Andamanese, 130- Nicobarese, 131- Shompen.

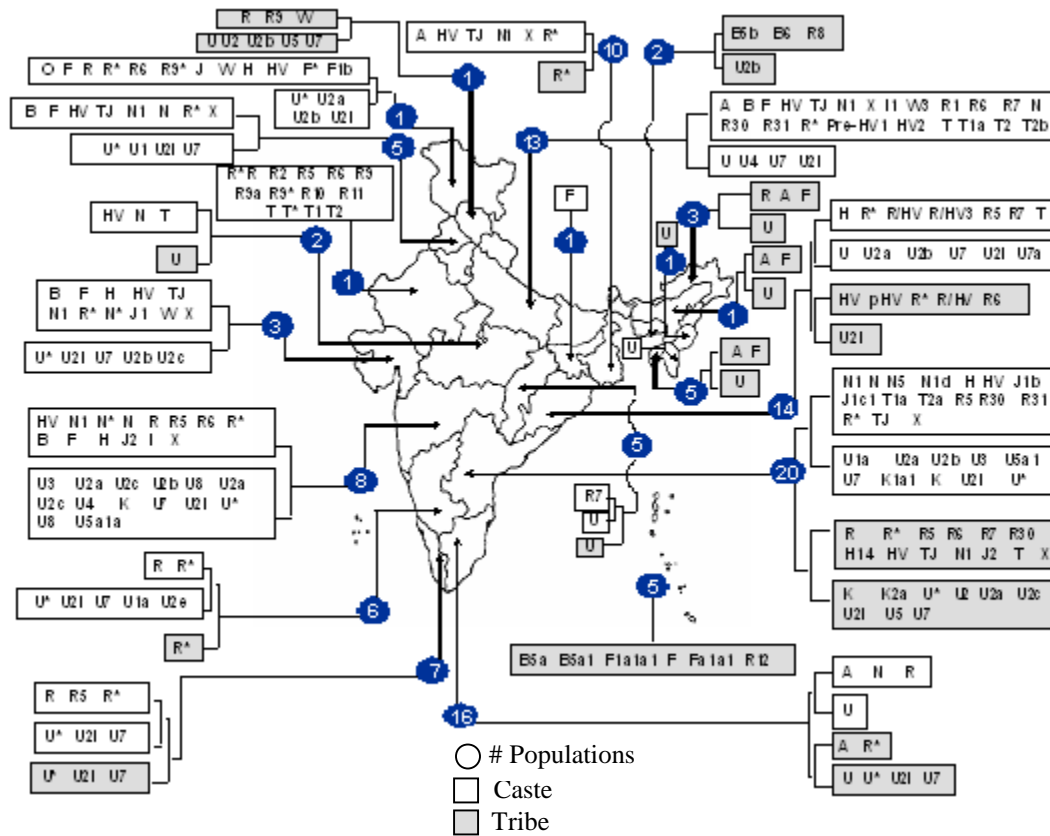


Fig. 2. Indian map showing region wise distribution of mitochondrial DNA diversity of macrohaplogroup N and its subgroup U among castes and tribes.

speaking tribal populations (Lodha (~18%), Santal (~10%)) possessed higher frequencies of this haplogroup. The tribal populations (Irula, Kota, Kurumba, Lodha, Santal, Munda, Muria, Tripuri) studied by Roychoudhury et al. (2000) were further analyzed for three additional RFLPs (*HpaI* np 3592, *AluI* np 7025 and *AluI* np 13262) and HVI region by Roychoudhury et al. (2001) and a frequency of about 10% (in the pooled sample) for haplogroup U was reported. Frequency of U was found to vary according to the linguistic affiliation of the populations, viz., with highest frequency observed among Austro-Asiatic (13.56%) followed by Dravidian (9.17%) and then by Tibeto-Burman (6.7%) speaking populations. Also, U2i and U1 (subclusters of U) were observed among the studied tribals with frequencies of 77.3% and 9.1%, respectively, of U.

Edwin et al. (2002) analyzed 160 individuals

from five Dravidian tribal populations (Kadar, Paniya, Irula, Kota and Kurumba) of southern India for ten RFLPs (*HaeIII* np 663, *AluI* np 5176, *DdeI* np 10394, *AluI* np 10397, *HinfI* np 12308, *HincII* np 13259, *HaeIII* np 16517, *HpaI* np 3592, *AluI* np 7025 and *AluI* np 13262) and one insertion/deletion polymorphism (COII/tRNA<sup>Lys</sup> intergenic 9-bp deletion). The study reported a pooled frequency of 14% for haplogroup U with highest frequency (~23%) among the Irula tribe of Tamil Nadu. Paniya, Kurumba and Kadar exhibited a frequency of around 7% for haplogroup U, whereas this haplogroup was completely absent among Kota tribe. Bamshad et al. (2001) performed HVI sequencing and RFLP (*HpaI* np 3592, *DdeI* np 10394, *AluI* np 10397, *AluI* np 13262, *BamHI* np 13366, *AluI* np 5176, *HaeIII* np 4830, *AluI* np 7025, *HinfI* np 12308, *AccI* np 14465, *Avall* np 8249, *AluI* np 10032, *BstOI* np 13704 and *HaeII*

np 9052) analysis of ~ 265 males from eight different caste populations of Andhra Pradesh that were ranked into upper (Niyogi and Vydiki

Brahmin, Kshatriya, Vysya [n = 80]), middle (Telega and Turpu Kapu, Yadava [n = 111]) and lower (Relli, Madiga and Mala [n = 74]) categories.

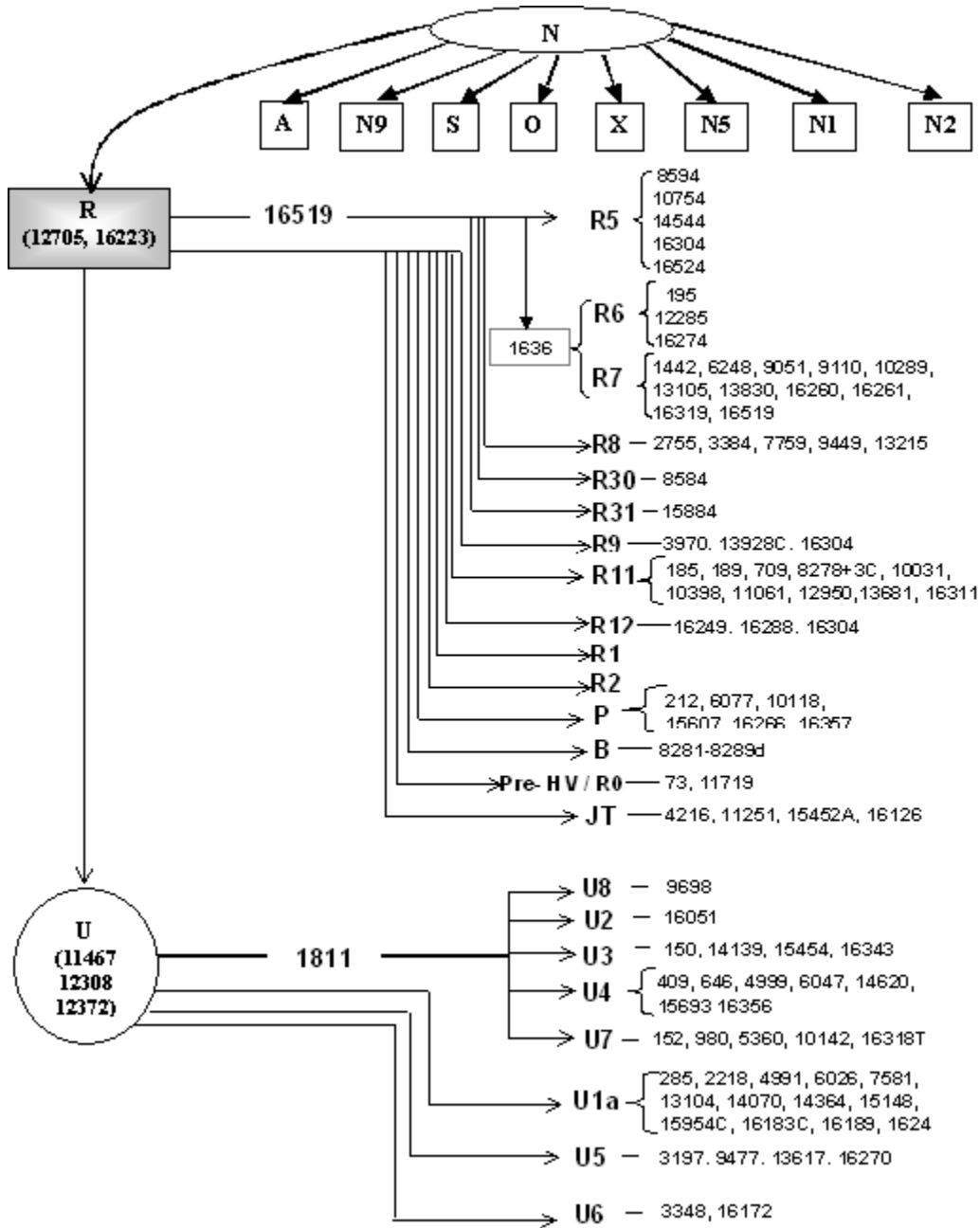


Fig. 3. Phylogenetic network of haplogroup R and its sub-haplogroups among Indian populations

Frequency of U2i haplotypes showed an increasing trend from lower (~ 6%) to upper (~ 17%) caste groups. Another study by Kivisild et al. (2003) based on the sequences of HVI, HVII, three segments (nps 1674 – 1880, 4761- 5260 and 8250 - 8710) of the coding region and also informative RFLP positions on Chenchu (n = 96) and Koya (n = 81) tribes of Andhra Pradesh, reported the presence of sub-haplogroups of U. Also 106 West Bengalis of different castes, 58 Konkanastha Brahmins from Bombay, 53 Gujaratis, 112 Punjabis of different caste ranks and 139 samples from Uttar Pradesh were studied and the obtained data was also compared with other published data from Kivisild et al. (1999), Roychoudhury et al. (2001), Mountain et al. (1995) and Bamshad et al. (2001). Haplogroup U was found to be rare in Chenchus and Koyas when compared with the other caste populations of India and U2i, U7 and U\* (U1, U3 – U6) were reported among the studied populations. Cordaux et al. (2003) analyzed 752 individuals belonging to 21 populations of the Indian subcontinent for HVI and compared with 219 previously published sequences from 10 Indian tribes (Kivisild et al. 1999; Roychoudhury et al. 2001) to understand the mitochondrial DNA sequence variation among Indian tribal populations. South Indian tribes (from Kerala, Karnataka and Tamil Nadu) exhibited a higher frequency (~ 15%) of haplogroup U followed by tribes from the east (Lodha, Munda and Santal [~ 14%]). Central (Andh, Pardhi, Thoti, Lambadi) and Northeast (Adi, Apatani, Nishi, Naga and Tipperah) tribes showed a low frequency, 6.3% and 1.7%, respectively, of haplogroup U.

To understand the peopling of India and the origins of the ethnic populations, Basu et al. (2003) studied 44 populations sampled from wide geographic regions and belonging to varied socio-cultural and linguistic categories for ten mtDNA RFLPs (*HaeIII* np 663, *AluI* np 5176, *DdeI* np 10394, *AluI* np 10397, *HinfI* np 12308, *HincII* np 13259, *HaeIII* np 16517, *HpaI* np 3592, *AluI* np 7025 and *AluI* np 13262), one Insertion/Deletion polymorphism (COII/tRNA<sup>Lys</sup> intergenic 9-bp deletion) and HVI sequences. U2i was the predominant sub-cluster among tribe and caste populations; with highest frequency (~ 90% of U) among the tribes. Tribes showed complete absence of U2e and low frequency of U1 and U7, whereas castes (upper, middle and lower) exhibited the presence of U2e in an increasing fashion as we move from lower to upper castes.

The frequency distribution of U1 was found to be similar in three ranked castes and that of U7 was different, with lower castes showing high predominance followed by upper and middle castes. HVI sequence analysis of 34 Gujarati samples from North-western India revealed equal predominance of U2a and U7 sub-haplogroups (8.8%) followed by U2b (5.9%) and U2c (2.9%). Other sub-haplogroups of U: U1, U2e, U3, U4, U5, U8b, U9 and K, were completely absent in the studied populations (Quintana-Murci et al. 2004). Metspalu et al. (2004) assayed 796 Indian samples covering a wide geographical area for HVI and some coding regions of mtDNA and compared the obtained data with the previously published data on Indian populations (Mountain et al. 1995; Bamshad et al. 1998; Kivisild et al. 1999; Kivisild et al. 2003; Roychoudhury et al. 2001; Cordaux et al. 2003; Quintana- Murci et al. 2004). The study reported the presence of haplogroup U in 15% of the castes and in 8% of the tribes studied. U2 (U2i: U2a, U2b and U2c) was found to be the predominant sub-haplogroup of U among the studied samples followed by another sub-haplogroup U7, constituting 12% of U in Gujarat and 9% in Punjab.

Palanichamy et al. (2004) performed the complete mtDNA sequencing of 75 samples from the populations of Reddy and Thogataveera from Andhra Pradesh, South India; Bhargava, Chaturvedi and other Brahmins from Uttar Pradesh (UP), North India; Rajbhansi from West Bengal, East India and Khasi from Meghalaya, Northeast India. This study reported the presence of the following sub-haplogroups of U: U1a, U2 (U2a, U2b, U2c, U2e), U3, U5 (U5a1), U7 and U8 (K: K1a1, K2a). Among the studied populations, U2a was observed in samples of UP Brahmin; U2b in samples of UP Brahmin, Reddy and Khasi populations; U2c in Reddy and Rajbhansi samples and U2e in Thogataveera and Reddy individuals. U3 and U5 (U5a1) was found in an individual from UP Brahmin population; U7 in Chaturvedi and other Brahmins of Uttar Pradesh and U8 in UP Brahmin (sub-haplogroup K1a1) and Thogataveera (sub-haplogroup K2a) populations. Barnabas et al. (2005) reported four sub-haplogroups of U: U1, U2i U4, and U7a, in his study consisting of 117 Indians (along with 27 Indians from Quintana-Murci et al. 1999) from different language groups inhabiting wide geographical regions. U1 was observed in individuals from Tamil Nadu and Kerala; U2i was

present in Tamil Nadu, Hindi speakers from Northern and Central regions and Marathi speakers from Maharashtra; U4 was exclusively present among the Telegu speakers of Andhra Pradesh and U7a was reported among Hindi speakers from Northern and Central regions, Sindhi speakers from Western region and Marathi speakers from Maharashtra. Based on HVI and coding region sequence variations, Gaikwad and Kashyap (2005) observed U2i, U2a, U2b, U2c, U3, U4, U5a1a, U7 and K sub-haplogroups of U in their study on Desasth Brahmins (n = 19), Chitpavan Brahmin (n= 20), Dhangar (n=19) and Maratha (n= 19) populations of Maharashtra, Western India. U accounted for 14.6% of the studied samples. Maratha samples possessed the sub-haplogroups U2i, U2a, U2b, U2c and U5a1a; Dhangars showed U2i, U2a, U2c, U4, U7 and K; Desasth Brahmins exhibited U2a, U2c and U7 and Chitpavan Brahmins displayed the presence of sub-haplogroup U3 only. Sahoo and Kashyap (2006) reported the mtDNA variation of seven populations from Orissa, Eastern India (Oriya Brahmin [n=20], Karan [n=17], Khandayat [n=16], Gope [n=16], Juang [n=20], Saora [n=19] and Paroja [n=21]) by employing HVI, HVII sequencing and RFLP (*AluI* np 10397, *DdeI* np 10394, *AluI* np 7025, *HinfI* np 12308, *DdeI* np 1715, *HaeIII* np 8994, *AluI* np 15606, *MseI* np 14766) analysis. Their analysis revealed the presence of U2a in one sample of Oriya Brahmin, Khandayat and Gope; U2b in a sample of Karan, Gope and in two samples of Khandayat; U2i in a sample of Paroja; U7 in a sample of Karan and U in a sample from Khandayat.

Thanseem et al. (2006) studied HVI, HVII and some selected coding regions of mtDNA among the three tribal populations of Andhra Pradesh - Pardhan (n = 193), Naikpod (n = 88) and Andh (n = 66) - and compared with the previous studies (Kivisild et al. 2003; Bamshad et al. 1998; Metspalu et al. 2004; Mountain et al. 1995; Cordaux et al. 2003). Among the different sub-haplogroups of U, Thanseem et al. (2006) observed the presence of U2, U2a and U5 in the studied samples. The highest frequency among the three sub-haplogroups of U was exhibited by U2 (4.61%) followed by U2a (3.17%) and then by U5 (2.31%). U2 was observed in all the three studied populations with higher frequencies in Andh and Pardhan (6.06% and 5.7%) than in Naikpod (1.14%). U2a was found in 8% of Naikpod samples and in around 2% of Pardhan samples whereas it

was completely absent in Andh population. U5 exhibited a frequency of about 3% in both Pardhan and Andh populations but was completely absent in Naikpod. Thangaraj et al. (2005b), in their study on 9-bp deletion/insertion polymorphism of the Indian populations reported the presence of U in a sample of Irula and Baduga (Tamil Nadu) and U7 in a sample of Lodha (West Bengal) and Yadava (Uttar Pradesh). Sub-haplogroup U2a was reported in a 9-bp deletion/insertion individual of Bathudi population from Orissa (Kumar et al. 2006).

### Other Sub-haplogroups of R

Kivisild et al. (1999) also observed other minor sub-haplogroups of R; H, J, T in the studied populations. Sub-haplogroups H and T exhibited equal frequencies (1.8%) with H being more frequent (3%) in north India (Uttar Pradesh: Tharu, Buksa, Mixed castes) than in south India (Andhra Pradesh: Telugus, Mixed castes) and T exhibited more or less similar frequency in both north (1%) and south (1.7%) India. J was reported only among the south Indian populations (0.4%) with complete absence among the north Indian populations. In their study on Andhra Pradesh caste populations, Bamshad et al. (2001) observed sub-haplogroups H, J and T, where upper castes (Niyogi and Vydiki Brahmin, Kshatriya, Vysya [n = 80]) showed the presence of only H (3.4%), middle castes (Telega and Turpu Kapu, Yadava [n = 111]) showed both J (0.9%) and T (2.7%) sub-haplogroups and lower castes (Relli, Madiga and Mala [n = 74]) displayed the presence of only T (1.4%). Haplogroup R lineages were reported in Chenchus (1%) and Koyas (31%) but this lineages did not further subdivide into western Eurasian specific (HV, TJ, R1) or eastern Eurasian specific (B, R9) branches (Kivisild et al. 2003). The coalescence time of haplogroup R was estimated to be 73,000 ± 20,900 years (Kivisild et al. 2003). To unravel the mtDNA diversity of Indian tribal populations, Cordaux et al. (2003) studied 752 individuals belonging to 21 populations of the subcontinent. Sub-haplogroups F, JT, H were reported; F being restricted only to northeast Indian tribes (15.5%), JT both to north (6.7%) and central (4.2%) Indian tribes and H to central (0.5%) and south (0.6%) Indian tribes. Mitochondrial DNA analysis of 34 samples of Gujaratis, northwestern India, by Quintana-Murci et al. (2004), revealed the presence of R\* (8.8%), H(5.9%), J1(2.9%) and

complete absence of other sub-haplogroups of R, viz., R1, R2, R5, R6, B, F, Pre-HV1, HV\*, HV1, HV2, V, J2, T\* and T1. Thangaraj et al. (2003) observed the presence of sub-haplogroup F in two individuals of Great Andamanese (n = 20) and seven individuals of Nicobarese (n = 14) from Andaman and Nicobar Islands.

Palanichamy et al. (2004) conducted a study on Reddy and Thogataveera from Andhra Pradesh, Chaturvedi, Bhargava and Brahmins from Uttar Pradesh, Rajbhansi from West Bengal, Khasi from Meghalaya based on mtDNA haplogroup N and reported the presence of R5 (Bhargava, Chaturvedi, Thogataveera and Reddy), R6 (Bhargava, Thogataveera and UP Brahmins), R7, R8, R30, R31, pre-HV ((pre-HV)1, HV2, H, pre-V: V2a) and JT (J1b1, J1c1, T1a, T2) sub-haplogroups of R. The study defined four novel sub-haplogroups of R: R7 (in Thogataveera and UP Brahmin), R8 (in Bhargava and Khasi), R30 (in Reddy, Thogataveera, UP Brahmins and Bhargava) and R31 (in UP Brahmins and Reddy). Sub-haplogroup HV was present in the samples of Thogataveera and Reddy of Andhra Pradesh and (pre-HV)1 and HV2 were found in UP Brahmins. Reddy samples also exhibited H along with Chaturvedi samples. V2a was found in Chaturvedi of Utttar Pradesh. J1b and J1c1 of J clade were reported in samples of Reddy population and J1b1 was observed in Chaturvedi sample. T1a of T clade was present in samples of UP Brahmin and Reddy populations; T2 in Bhargava and UP Brahmins; T2a, T2b in Reddy and UP Brahmins, respectively. Based on 796 mtDNA Indian samples, Metspalu et al. (2004) observed R5 to be the second most frequent sub-haplogroup among the caste populations. In spite of being absent in Austro-Asiatic tribal groups R5 was distributed across other language groups of the subcontinent. The coalescence time was estimated to be  $66,100 \pm 22,000$  years. Sub-haplogroup R2 was rare among Indian samples and the age of this sub-clade was calculated to be  $11,400 \pm 9,000$  ybp. Barnabas et al. (2005), in their study on 117 samples from different language groups of India, reported haplogroup R in Tamil, Kannada, Malayalam and Marathi speaking populations of India, sub-haplogroup R5 in Malayalam and Marathi speaking populations, T in Hindi and Marwari speaking populations of northern and central region, HV in Urdu speaking populations and F in Bihari community from Bihar of eastern India. Thangaraj

et al. (2005a) classified the Nicobarese B5a sequences into B5a1a defined by three coding region mutations and the coalescence time of  $1.7 \pm 1.7$  kya was determined for B5a1a. Nicobarese F sequence was categorized into F1a1a1 and its age was estimated to be  $16.7 \pm 5.6$  kya. Gaikwad and Kashyap (2005) reported the following sub-haplogroups of R: R\*, R5, R6, HV, H, J2 in the caste populations of Maharashtra (Desasth Brahmin, Chitpavan Brahmin, Dhargar and Maratha) castes of Maharashtra. While Dhangers showed R\*, R5 and R6, Desasth Brahmin R\*, HV, and J2, and Chitpavan Brahmin R\*, R5, HV, and H, Maratha showed only sub-haplogroup H. Sahoo and Kashyap (2006) reported the mtDNA variation among seven populations of Orissa, Eastern India: Oriya Brahmin (n=20), Karan (n=17), Khandayat (n=16), Gope (n=16), Juang (n=20), Saora (n=19) and Paroja (n=21). Sub-haplogroups H, HV, Pre-HV, R\*, R5, R6, R7 and T were observed among these populations. Sub-haplogroups pre-HV and HV were present in a sample of Paroja, H in a sample of Khandayat and R\* was found in 13 samples (Oriya Brahmin [2], Karan [2], Khandayat [2], Gope [1], Saora [1] and Paroja [5]). Beside these sub-haplogorups of R, R5 was found in 2 Gope samples, R6 was reported in 3 Juang samples, T was found in only a sample of Oriya Brahmin and R7 was observed in both Karan (2) and Gope (1) samples.

Thanseem et al. (2006) reported the presence of R\*, R5, R6, R7, H14, J2, T in Andh, Naikpod and Pardhan tribal populations of Andhra Pradesh, south India, where R was observed the maximum in Naikpod (~ 11%) followed by Pardhan (~ 6%) and then by Andh (~ 2%), R5 was found in more or less similar frequency (~ 12%) in Pardhan and Andh populations and at comparatively lower frequency (~ 5%) in Naikpod and R6 and R7 were present at lower frequencies in Pardhan (1.04%) and Andh (3.03% and 1.52%) but absent in Naikpod. H14 and J2 were reported in Naikpod (3.41%) and Pardhan (3.11%), respectively and sub-haplogroup T was present in Pardhan (0.52%) and Andh (1.52%) but completely absent in Naikpod. Trivedi et al. (2006) observed the sub-haplogroups B5a and R12 in their molecular genetic study on 29 Shompen individuals from Great Nicobar Islands. B5a was reported in 19 individuals and their newly defined sub-haplogroup R12 was found in 10 Shompen individuals. A 9-bp deletion sample from Nishi (Arunachal Pradesh), Kathodi and Koli (Gujarat)



of sub-haplogroup R and a Kanwar (Chattisgarh) sample of sub-haplogroup R7 was reported (Thangaraj et al. 2005b). Sub-haplogroups H and T were present in a Gond sample of Madhya Pradesh and in a sample of Rohidas of Uttar Pradesh, respectively. Sub-haplogroup B was reported in a random sample from West Bengal, B4 in a sample of Tiwari (Uttar Pradesh) and Bengali (West Bengal), B5a, B5a1 and B5b in a sample of Rajput (Uttar Pradesh), 22 samples of Nicobarese (Nicobar Islands) and a sample of Gond (Madhya Pradesh), respectively (Thangaraj et al. 2005b). Kumar et al. (2006) in their study on 9-bp Deletion/insertion polymorphism of 31 tribal population of India observed sub-haplogroup R and R7 in Bhuiyan samples of Orissa, B5b in a Garo sample and B6 in 6 samples of Khasi (5 Maram and 1 Bhoi) from Meghalaya.

#### OTHER MINOR HAPLOGROUPS OF N

Kivisild et al. (1999) observed other sub-haplogroups of N: I, W and X in their study among the Tharu, Buksa, Mixed castes of Uttar Pradesh and Telugu mixed castes from Andhra Pradesh. North Indian samples showed the presence of only sub-haplogroup I (2%) and the South Indian samples exhibited the presence of only sub-haplogroup W (0.4%) among the reported sub-haplogroups I, W and X. The overall frequency of I, W and X in India was calculated to be 0.7%, 2.2% and 0.2%, respectively. Along with the above reported west Eurasian sub-haplogroups, A and O were also observed at lower frequencies in their samples. Roychoudhury et al. (2000) observed East-Asian specific sub-haplogroup A in Muslims (14.3%) and Rajputs (5.9%) of Uttar Pradesh, Mahishya (9.1%) of West Bengal and Tipperah (8.9%) of Tripura. Sub-haplogroup A also represented 3 to 4% of Brahmins of Uttar Pradesh, Iyer and Kurumba of Tamil Nadu. Sub-haplogroup A was again reported by Roychoudhury et al. (2001) in Tipperah of Tripura (8.9%) and Kurumba (3.3%) of Tamil Nadu and by in Kurumba (3.3%) of Tamil Nadu (Edwin et al. 2002). The study on ranked caste populations of Andhra Pradesh by Bamshad et al. (2001) reported the presence of only sub-haplogroup W at low frequencies (1.7%) in upper caste populations (Niyogi and Vydiki Brahmin, Kshatriya, Vysya [n = 80]) and its absence in middle (Telega and Turpu Kapu, Yadava [n = 111]) and lower (Relli, Madiga and

Mala [n = 74]) caste populations. Sub-haplogroups I and X were completely absent among the populations of three caste ranks. Cordaux et al. (2003) reported the presence of sub-haplogroup A (16.7%) only among the tribes of north-east India (Adi, Apatani, Nishi, Naga, Tipperah) and sub-haplogroup W (1.6%) only among the Central Indian tribes (Andh, Pardhi, Thoti, Lambadi). Sub-haplogroup W along with U7 and R2 constitute one third of the west Eurasian specific mtDNAs found in India. W was reported to be the most frequent sub-haplogroup of N after R and its frequency reaches up to 5% in northwestern states (Gujarat, Punjab and Kashmir), whereas its frequency in other parts of India is very low (0 – 0.9%) thereby forming a significant spatial cline (Metspalu et al. 2004). The coalescence time of sub-haplogroup W was calculated to be  $37,900 \pm 11,100$  years (Metspalu et al. 2004). Palanichamy et al. (2004) reported sub-haplogroup N5 and N1d in Reddy, I1 and W3 in Brahmins and W1c and W in Chaturvedi of Uttar Pradesh. Quintana-Murci et al. (2004) observed sub-haplogroup W at a frequency of 8.8% in 34 Gujarati samples from Northwestern India, Gujarat and N\* was also reported (2.9%) in these samples and Barnabas et al. (2005) found sub-haplogroup I in a Marathi speaking sample of Maharashtra and this study reported N in Punjabis, Tamil, Telegu, Sindhi, Hindi and Marathi-speaking samples. In a study on caste populations of Maharashtra (Desasth-Brahmin, Chitpavan Brahmin, Dhangar and Maratha), Gaikwad and Kashyap (2005) reported sub-haplogroup W in a sample of Maratha and Desasth-Brahmin and their study also categorized a sample of Desasth-Brahmin under N\*. Sahoo and Kashyap (2006) observed sub-haplogroups W and X in their study on seven population groups of Orissa, eastern India. W was reported in 6 samples; Oriya Brahmin (1), Karan (2), Khandayat (1) and Juang (2) and X were present in a sample of Saora. Study on Pardhan, Andh and Naikpod populations of Andhra Pradesh by Thanseem et al. (2006) revealed the presence of only sub-haplogroup W in Andh (1.52%), whereas the other two studied populations, Pardhan and Naikpod, lacked the sub-haplogroup.

#### Patterns of Variation

Overall, the mtDNA haplogroup distribution of N and its subhaplogroups studied among

about 130 castes and tribes from different regions show wide diversity. The number of studied castes and tribes highly vary between regions from 4 castes in central region to 25 castes in southern region and 3 tribes in central region a maximum of 24 in southern region. Despite the disparity in the number of castes and tribes studied in different regions certain trends of variation in haplotypes can be observed for subhaplogroup R and its subhaplogroup U. For example, the western region exhibits a high of 39 R subhaplogroups among 12 castes studied whereas central region shows four subhaplogroup among four castes studied. In case of southern region 25 castes studied show 18 different R subhaplogroups. Among tribes southern region shows as many as 17 R-subhaplogroups among 24 populations studied and the tribes from island regions show 6 haplogroups. In central region three tribes studied hardly show any R-haplogroup diversity.

In case of U-subhaplogroup distribution in different regions, western region shows the highest of 18 different haplogroups among 12 castes and 19 haplogroups have been observed among 25 castes studied in southern region. It was absent among 4 castes of central region. Among tribes, northern region one tribe showed 5 subhaplogroups of U, whereas the northeastern region shows more homogeneity among the groups studied; 13 tribes exhibit only 5 different U-subhaplogroups. Equally all the 3 tribes in central region show single subhaplogroup. However southern region shows higher

diversity than the other regions, it shows 16 different haplogroups of U observed among 24 tribes studied.

Overall, between three different regions, where it is possible to compare, the average number haplogroups of R and U show almost an equal frequency among the castes and tribes. However the total observed frequency of the haplogroups vary between castes and tribes, high in southern region than in eastern and northern regions (Fig. 4). A similar trend is observed among the linguistic groups; the average number of different haplogroups of R and U show about equal frequency between the three regions, however the total observed frequency vary; higher in southern region than the other two regions (Fig. 5). Overall, the distribution of different individuals belonging to different subhaplogroups of both R and U show significant association between castes, tribes and regions ( $\chi^2(4,1) = 64.56$ ;  $p < 0.001$ ). In case of U, subhaplogroup, the chi-square test also shows significant association between castes, tribes and between regions ( $\chi^2(4,1) = 28.67$ ;  $p < 0.001$ ).

When compared to more than four thousand and odd diverse ethnic communities (4635), the populations studied for mtDNA diversity in India is small (2.9%). Possibly more mtDNA diversity studies on diverse ethnic populations of India will provide a better understanding of the maternal genetic history, and clear phylogenetic position of Indian populations besides revealing more patterns that could be related to geography, language and ethnicity.

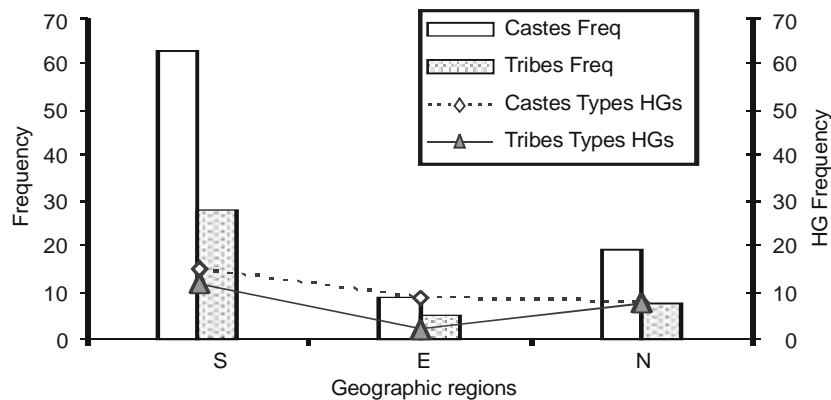


Fig. 4. Region wise distribution of number of types of N subhaplogroups (Y2 axis) and its total frequency among castes and tribes.

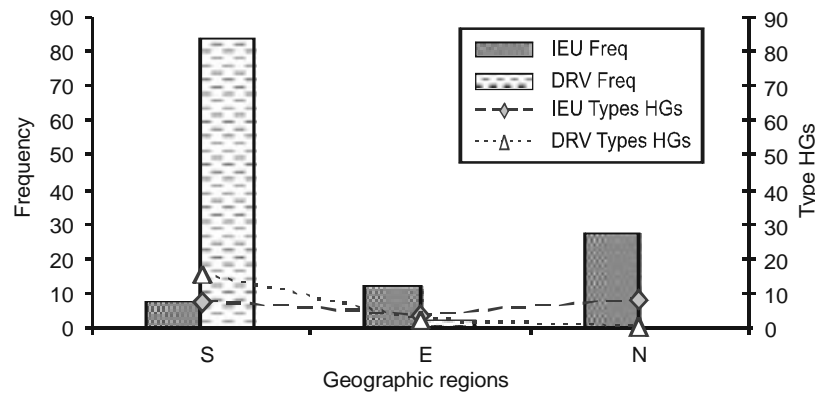


Fig. 5. Region wise distribution of number of types of N subhaplogroups (Y2 axis) and its total frequency among Dravidian (DRV) and Indo-European (IEU) Linguistic groups studied.

## CONCLUSIONS

Mitochondrial DNA macrohaplogroup N among Indian populations shows wide diversity of haplogroup R and its sub-haplogroups U with characteristic high frequency of Indian specific variants, especially R5 – R7 and U2i (U2a, U2b, U2c) possibly suggesting their autochthonous status. Possibly, N5 is also of Indian origin as ancient as R. The distribution of R and U appear to be least affiliated to language, ethnicity or geography.

## REFERENCES

- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG 1981. Sequence and organization of the human mitochondrial genome. *Nature*, **290**: 457-465.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N 1999. Reanalysis and revisions of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*, **23**: 147.
- Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, Wallace DC 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics*, **130**: 139-152.
- 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.
- 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, 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.
- Barnabas S, Shouche Y, Suresh CG 2005. High-resolution mtDNA studies of the Indian population: Implications for paleolithic settlement of the Indian subcontinent. *Ann Hum Genet*, **70**: 42-58.
- Brown WM, 1980. Polymorphism in mitochondrial DNA of human as revealed by restrictive endonuclease analysis. *Proc Nat Acad Sc, USA*, **77**: 3605-3609.
- Cann RL, Stoneking M, Wilson AC 1987. Mitochondrial DNA and human evolution. *Nature*, **325**: 31-36.
- Cann RL, 2001. Genetic clues to dispersal of human populations: Retracing the past from the present. *Science*, **291**: 1742-1748.
- Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC 1995. Analysis of mtDNA variation in African population reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet*, **57**: 133-149.
- 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.
- Danaro M, Blanc H, Johnson MJ, Chen KH, Wilmsen E, Cavalli Sforza LL, Wallace DC 1981. Ethnic variation in HpaI endonuclease cleavage patterns of human mitochondrial DNA. *Proc Nat Acad Sc, USA*, **78**: 5768-5772.
- Edwin D, Vishwanathan H, Roy S, Usha Rani MV, Majumder PP 2002. Mitochondrial DNA diversity among five tribal populations of southern India. *Curr Sci*, **83**: 158-162.
- Finnala S, Lehtonen MS, Majanaa K 2001. Phylogenetic network for European mtDNA. *Am J Hum Genet*, **68**: 1475-1484.
- Gaikwad S, Kashyap VK 2005. Molecular insight into the genesis of ranked caste populations of western India based upon polymorphisms across non-recombinant and recombinant regions in genome. *BMC Genome Biol*, **6**: 10.

- Ingman M, Kaessmann H, Paabo S, Gyllensten U 2000. Mitochondrial genome variation and the origin of modern humans. *Nature*, **408**: 708-713.
- 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, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R 2003. The genetic heritage of earliest settlers persist in both the Indian tribal and caste populations. *Am J Hum Genet*, **72**: 313-332.
- Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am J Hum Genet*, **75**: 752-770.
- Kivisild T, Shen P, Wall EP, Do B, Sung R, Davis K, Passarino D, Goppa A, de Knijff P, Feldman M, Cavalli-Sforza LL, Oefner PJ 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics*, **172**: 373-387.
- Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, Zhang YP 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet*, **73**: 671-676.
- Kumar V, Langsateh BT, Biswas S, Babau JP, Rao TN, Thangaraj K, Reddy AJ, 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.
- Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Deehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt HJ, Oppenheimer S, Torroni A, Richards M 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science*, **308**: 1034-1036.
- Maca-Mayer N, Gonzatez AM, Larruga JM, Flores C, Cabrera VC 2001. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics*, **2**: 13.
- Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MTP, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R 2004. Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet*, **5**: 26.
- Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace D 2003. Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci*, **100**: 171-176.
- 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.
- 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.
- Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS 1999. Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet*, **23**: 437-441.
- 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.
- Roychoudhury S, Roy S, Dey B, Chakraborty M, Roy M, Roy B, Ramesh A, Prabhakaran M, Usha Rani MV, Vishwanathan H, Mitra M, Sil SK, Majumder PP 2000. Fundamental genomic unity of ethnic India is revealed by analysis of mitochondrial DNA. *Curr Sci*, **79**: 1182-1192.
- 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.
- Sahoo S, Kashyap VK 2006. Phylogeography of Mitochondrial DNA and Y-Chromosome haplogroups reveal asymmetric gene flow in populations of Eastern India. *Am J Phy Anthropol*, **131**: 84-97.
- Salas A, Richaqrds M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, Macaulay V and Carracedo A 2002. The making of the African mtDNA landscape. *Am J Hum Genet*, **71**: 1082-1111.
- Thangaraj K, Singh L, Reddy AJ, Rao R, Sehgal CS, Underhill PA, Pierson M, Frame I G, Hagelberg E 2003. Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr Biol*, **13**: 86-93.
- Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar A, Singh L 2005a. Reconstructing the origin of Andaman Islanders. *Science*, **308**: 996.
- Thangaraj K, Sridhar V, Kivisild T, Reddy AG, Chaubey G, Singh VK, Kaur S, Agarawal P, Rai A, Gupta J, Basu Mallick C, Kumar NP, Velavan T, Suganthan R, Udaykumar R, Kumar R, Mishra R, Khan A, Annapurna C, Singh L 2005b. Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. *Hum Genet*, **116**: 507-517.
- Thanseem I, Thangaraj K, Chaubey G, Singh VK, Bhaskar VKSL, Reddy BM, Reddy A G, Singh L 2006. Genetic affinities among the lower caste and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC Genetics*, **7**: 42.
- 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.
- Watson E, Forster P, Richards M, Bandelt HJ 1997. Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet*, **61**: 691-704.