

REVIEW ARTICLE

Phylogeographic distribution of mitochondrial DNA macrohaplogroup M in India

SUVENDU MAJI, S. KRITHIKA and T. S. VASULU*

Biological Anthropology Unit, Indian Statistical Institute, Kolkata 700 108, India

Abstract

Indian subcontinent harbours both the human mtDNA macrohaplogroups M and N, of which M is the most prevalent. In this study, we discuss the overall distribution of the various haplogroups and sub-haplogroups of M among the different castes and tribes to understand their diverse pattern with respect to geographical location and linguistic affiliation of the populations. An overview of about 170 studied populations, belonging to four distinct linguistic families and inhabiting different geographic zones, revealed wide diversity of about 22 major haplogroups of M. The tribal populations belonging to the same linguistic family but inhabiting different geographical regions (Dravidian and Austro-Asiatic speakers) exhibited differences in their haplogroup diversity. The northern and southern region castes showed greater diversity than the castes of other regions.

[Maji S., Krithika S. and Vasulu T. S. 2009 Phylogeographic distribution of mitochondrial DNA macrohaplogroup M in India. *J. Genet.* **88**, 127–139]

Introduction

The advancement in molecular biology has broadened our perspective on the origin and evolution of anatomically modern humans, thereby furnishing substantial evidence in support of the existing archeological, anthropological and other biological records (Cann *et al.* 1987; Excoffier and Langaney 1989; Ingman *et al.* 2000; Macaulay *et al.* 2005). In recent years, high-resolution-genetic markers (*viz.* autosomal, mtDNA and Y-chromosomal) have been utilized to test various hypotheses related to past genetic history and evolution of man (Stringer and Andrews 1988; Vigilant *et al.* 1991; Rogers and Jorde 1995; Hammer *et al.* 1998; Templeton 2002; Hebsgaard *et al.* 2007). In particular, the complete sequencing of the 16,569 base pairs of human mitochondrial genome reported by Anderson *et al.* (1981) widened the employment of mitochondrial DNA (mtDNA) as a genetic marker to investigate one of the major contentious issues, the time of origin and dispersal of man out of Africa to other continents (Cann *et al.* 1987; Ingman *et al.* 2000; Macaulay *et al.* 2005). Also, the maternal inheritance pattern of mtDNA enables us to disentangle the maternal genetic history of human populations. In addition, several other features of

mitochondrial genome, such as its high copy number, near-absence of genetic recombination and higher mutation rate (three to five times greater than nuclear DNA), make it a potent molecular genetic tool.

Initially, studies on human mtDNA were based on restriction fragment length polymorphisms (RFLPs) of either genomic DNA or mtDNA (Brown 1980; Denaro *et al.* 1981; Johnson *et al.* 1983; Cann *et al.* 1987; Scozzari *et al.* 1988; Excoffier and Langaney 1989), but with the advent of polymerase chain reaction (PCR) and sequencing techniques, RFLP analysis of mtDNA PCR products and sequence analysis of the highly polymorphic noncoding regions, hypervariable regions I and II, took over (Vigilant *et al.* 1991; Hedges *et al.* 1992; Maddison *et al.* 1992; Templeton 1992; Torroni *et al.* 1996, 1998; Watson *et al.* 1997; Macaulay *et al.* 1999; Quintana-Murci *et al.* 1999). Later, however, certain disadvantages of hypervariable regions, like back mutation, parallel substitution and mutation rate heterogeneity led researchers to sequence the whole genome using high-throughput sequencing, wherein many informative polymorphic sites are present, the numbers of back and parallel mutations are almost zero, and the problem of mutation rate heterogeneity is absent (Finnilä *et al.* 2000; Ingman *et al.* 2000; Elson *et al.* 2001; Finnilä and Majamaa

*For correspondence. E-mail: vasulu@gmail.com; vasulu@isical.ac.in.

Keywords. mtDNA; macrohaplogroup M; haplogroups; castes and tribes; linguistic families.

2001). Based on the mutation sites observed in the mitochondrial genome in relation to the reference sequence (Anderson *et al.* 1981; Andrews *et al.* 1999), individuals were categorized into specific monophyletic clades or haplogroups, where haplogroups represent related groups of sequences defined by shared mutations. The mtDNA haplogroups mainly fall into three macrohaplogroups, designated as L, M and N, distributions of which are geographically distinct.

L is the oldest macrohaplogroup, restricted to Africans, especially sub-Saharan African populations, and consists of haplogroups L0, L1, L2, L3, L4, L5, L6 and L7, of which L3 radiated out of Africa in the form of macrohaplogroups M and N around ~60,000 ybp (Quintana-Murci *et al.* 1999; Mishmar *et al.* 2003) and entered South Asia or Indian subcontinent, and later drifted to Southeast Asia and Australia. Recent studies suggest that Indian subcontinent had played a major role in the early genetic history of mankind, which can possibly be understood by M and N distribution among different populations of India (Kivisild *et al.* 1999; Metspalu *et al.* 2004; Palanichamy *et al.* 2004; Thangaraj *et al.* 2006).

Macrohaplogroup M (489-10400-14783-15043), excluding M1 which is east African, is distributed among most south, east and north Asians, Amerindians (containing a minority of north and central Amerindians and a majority of south Amerindians), and many central Asians and Melanesians. On the other hand, macrohaplogroup N (8701-9540-10398-10873-15301) has been reported among nearly all Europeans, west Asians, North Africans and Australian aborigines as well as among east Asians, South Asians, Amerindians (containing a majority of north and central Amerindians and a minority of south Amerindians), and Polynesians (Schurr *et al.* 1990; Ballinger *et al.* 1992; Chen *et al.* 1995; Torroni *et al.* 1996, 2001; Finnilä *et al.* 2001; Ingman and Gyllensten 2001; Maca-Meyer *et al.* 2001, 2003; Salas *et al.* 2002; Kong *et al.* 2003; Mishmar *et al.* 2003).

Each of the mtDNA macrohaplogroups is widely diversified in different regions of the globe with characteristic patterns of its own that are specific and unique to the region. Africa harbours the largest diversity of mtDNA haplogroups, comprising primarily of L and its haplogroups. Of the other parts of the globe, the next highest mtDNA diversity (in the form of M and N distribution) is traced to South Asia, in particular to Indian subcontinent (Basu *et al.* 2003; Metspalu *et al.* 2004; Palanichamy *et al.* 2004; Quintana-Murci *et al.* 2004; Sun *et al.* 2006; Thangaraj *et al.* 2006; Chaubey *et al.* 2007). This could be attributed to the major role that the Indian subcontinent had played during early period of human evolution after dispersal of man out of Africa. Among the two macrohaplogroups M and N, M is more prevalent and shows a wide diversity with deep-rooted lineages and region-specific unique types that are hardly reported in other regions of the world. The wide mtDNA diversity observed in Indian subcontinent is yet another dimension of biological and cultural diversity, that has been a unique feature of the Indian populations (Ratnagar 1995; Thapar 1995).

In this context, understanding the vast diversity of Indian populations is of much significance in unraveling the region specific diversity that could help in obtaining clarity about the past genetic history and human migration from Indian subcontinent to other regions, especially southeast Asia, Oceania and Australia. So far, mtDNA studies on Indian populations have been sporadic and these have reported many haplogroups and sub-haplogroups of M in different populations. However there is a dearth of a comprehensive overall view of vast diversity and distribution of M haplogroups and its sub-haplogroups, and their variation across linguistic, geographic and ethnic boundaries of the Indian populations. Such a comprehensive overview would aid in obtaining a glimpse of the entire M distribution in the subcontinent, which would enhance our understanding of the phylogeographic landscape of the populations and in turn would facilitate the formulation of future studies on mtDNA in South Asia and elsewhere. The present study attempts to comprehend the distribution of various haplogroups and sub-haplogroups of macrohaplogroup M and their association with the geography and linguistic affiliations of the population of India. The details of the populations studied in India for the mtDNA polymorphisms and their geographical affiliation are summarized in table 1 and figure 1.

Macrohaplogroup M in Indian populations

Studies on mtDNA variation in India (Kivisild *et al.* 1999; Roychoudhury *et al.* 2000; Rajkumar *et al.* 2005) have reported the presence of macrohaplogroup M in a majority of the individuals (~60%). The frequency of M was observed to be 58% among the caste and 72% among the tribal populations (Bamshad *et al.* 1998, 2001; Kivisild *et al.* 1999, 2003; Roychoudhury *et al.* 2000, 2001; Basu *et al.* 2003; Metspalu *et al.* 2004). The origin of M has been controversial and several hypotheses have been put forward to explain its origin. Some authors have suggested a southwest Asian origin for macrohaplogroup M, followed by a back migration to Africa (Kivisild *et al.* 2003) while others support its African ancestry (Quintana-Murci *et al.* 1999).

Among the defined haplogroups of M: M2, M3, M4, M5, M6, M18, M25, M30, M31, M32, M33, M34, M35, M36, M37, M38, M39, M40, M41, M48, M49 and M50 are the India-specific lineages. The India-specific detailed phylogeny of macrohaplogroup M is shown in figure 2.

Haplogroup M2

Haplogroup M2, defined by the motif 447G-1780-8502-11083-15670-16274-16319, consists of sub-haplogroups M2a (5252-7961-8396-9758-12810-16270-16319) and M2b (16274-16319-16357) (Kivisild *et al.* 2003; Rajkumar *et al.* 2005; Sun *et al.* 2006; Thangaraj *et al.* 2006). M2 has been characterized to be the most ancient haplogroup in the subcontinent and approximately one tenth of the Indian haplogroup M fall into M2 (Kivisild *et al.* 2003).

mtDNA diversity among Indian populations

Table 1. Number of populations studied, in different geographical regions of India for mtDNA polymorphisms, with special reference to macrohaplogroup M.

Geographic region	No. of populations studied	References	Denotation in figure 1
North			
Jammu and Kashmir	1	Kivislid <i>et al.</i> 1999.	1
Himachal Pradesh	1	Metspalu <i>et al.</i> 2004.	2
Punjab	6	Kivisild <i>et al.</i> 1999, 2003; Basu <i>et al.</i> 2003; Metspalu <i>et al.</i> 2004; Thanseem <i>et al.</i> 2006.	3–8
Uttar Pradesh	10	Kivisild <i>et al.</i> 1999, 2003; Roychoudhury <i>et al.</i> 2000; Basu <i>et al.</i> 2003; Cordaux <i>et al.</i> 2003; Metspalu <i>et al.</i> 2004; Sun <i>et al.</i> 2006; Thangaraj <i>et al.</i> 2005b; Kumar <i>et al.</i> 2006; Thanseem <i>et al.</i> 2006.	9–18
Northeast			
Arunachal Pradesh	3	Cordaux <i>et al.</i> 2003, 2004; Thangaraj <i>et al.</i> 2005b.	19–21
Assam	1	Cordaux <i>et al.</i> 2003, 2004.	22
Manipur	1	Basu <i>et al.</i> 2003.	23
Mizoram	1	Basu <i>et al.</i> 2003.	24
Tripura	5	Roychoudhury <i>et al.</i> 2000, 2001. Basu <i>et al.</i> 2003; Cordaux <i>et al.</i> 2003, 2004;	25–29
Meghalaya	10	Sun <i>et al.</i> 2006; Reddy <i>et al.</i> 2007.	30–39
East			
West Bengal	13	Roychoudhury <i>et al.</i> 2000, 2001; Basu <i>et al.</i> 2003; Cordaux <i>et al.</i> 2003; Metspalu <i>et al.</i> 2004; Thanseem <i>et al.</i> 2006; Sun <i>et al.</i> 2006; Rajkumar <i>et al.</i> 2005; Kumar <i>et al.</i> 2006.	40–52
Bihar	5	Rajkumar <i>et al.</i> 2005; Thangaraj <i>et al.</i> 2005b, 2006;	53–57
Orissa	15	Roychoudhury <i>et al.</i> 2000, 2001; Basu <i>et al.</i> 2003; Cordaux <i>et al.</i> 2003; Rajkumar <i>et al.</i> 2005; Kumar <i>et al.</i> 2006; Sahoo and Kashyap 2006; Thangaraj <i>et al.</i> 2006.	58–72
Jharkhand	7	Thangaraj <i>et al.</i> 2005b, 2006; Rajkumar <i>et al.</i> 2005; Kumar <i>et al.</i> 2006;	73–79
West			
Gujarat	3	Metspalu <i>et al.</i> 2004; Thangaraj <i>et al.</i> 2006.	80–82
Maharashtra	9	Basu <i>et al.</i> 2003; Kivisild <i>et al.</i> 2003; Metspalu <i>et al.</i> 2004; Rajkumar <i>et al.</i> 2005; Gaikwad and Kashyap 2005; Thanseem <i>et al.</i> 2006.	83–91
Rajasthan	1	Metspalu <i>et al.</i> 2004; Thanseem <i>et al.</i> 2006.	92
Central			
Madhya Pradesh	2	Thangaraj <i>et al.</i> 2005b; Watkins <i>et al.</i> 1999.	93–94
Chattisgarh	7	Roychoudhury <i>et al.</i> 2000, 2001; Basu <i>et al.</i> 2003; Thangaraj <i>et al.</i> 2005b; Kumar <i>et al.</i> 2006.	95–101
South			
Kerala	12	Mountain <i>et al.</i> 1995; Edwin <i>et al.</i> 2002; Cordaux <i>et al.</i> 2003; Basu <i>et al.</i> 2003; Metspalu <i>et al.</i> 2004; Thanseem <i>et al.</i> 2006; Thangaraj <i>et al.</i> 2006.	102–113
Karnataka	12	Mountain <i>et al.</i> 1995; Cordaux <i>et al.</i> 2003; Rajkumar <i>et al.</i> 2005; Thanseem <i>et al.</i> 2006.	114–125
Tamil Nadu	19	Watkins <i>et al.</i> 1999; Roychoudhury <i>et al.</i> 2000, 2001; Edwin <i>et al.</i> 2002; Basu <i>et al.</i> 2003; Cordaux <i>et al.</i> 2003; Thangaraj <i>et al.</i> 2005b;	126–144

Table 1 (contd.)

Geographic region	No. of populations studied	References	Denotation in figure 1
Andhra Pradesh	21	Thanseem <i>et al.</i> 2006; Thangaraj <i>et al.</i> 2006. Bamshad <i>et al.</i> 1998, 2001; Kivisild <i>et al.</i> 1999, 2003; Cordaux <i>et al.</i> 2003; Sun <i>et al.</i> 2006; Rajkumar <i>et al.</i> 2005; Thangaraj <i>et al.</i> 2005b, 2006; Thanseem <i>et al.</i> 2006;	145–165
Andaman and Nicobar Islands	5	Thangaraj <i>et al.</i> 2003, 2005a, b; Endicott <i>et al.</i> 2003; Thangaraj <i>et al.</i> 2006.	166–170



Figure 1. Map of India showing the location of the populations studied till date for mtDNA polymorphisms with special reference to macrohaplogroup M. 1, Mixed castes; 2, Kanet; 3, Lobana; 4, Jat Sikh; 5, Punjab Brahmins; 6, Scheduled Castes; 7, Mixed Caste; 8, Khatri; 9, Kori; 10, Tharu; 11, Bhoksha/ Buksa; 12, Mixed Caste; 13, Chamar; 14, Bhargava; 15, Chaturvedi; 16, Rajput; 17, Yadava; 18, UP Brahmins; 19, Adi; 20, Apatani; 21, Nishi; 22 Naga; 23, Meitei; 24, Mizo; 25, Chakma; 26, Jamatia; 27, Mog; 28, Riang; 29, Tipperah / Tripuri; 30, Khasi; 31, Bhoi; 32, Maram; 33, Lyngnam; 34, Nongtrai; 35, War Jaintia; 36, War Khasi; 37, Pnar; 38, Khyntiam; 39, Garo; 40, Bagdi; 41, Brahmins; 42, Lodha; 43, Mahishya; 44, Toto; 45, Mixed castes; 46, Kurmi; 47, Maheli; 48, Rajbansi; 49, Santal; 50, Bhumij; 51, Paharia; 52, Kharia; 53, Yadava; 54, Kurmi; 55, Baniya; 56, Oraon; 57, Rajput; 58, Santal; 59, Kharia; 60, Agharia; 61, Gaud; 62, Brahmins; 63, Karan; 64, Khandayat; 65, Gope; 66, Juang; 67, Paroja; 68, Saora; 69, Gadaba; 70, Tanti; 71, Bathudi; 72, Munda; 73, Munda; 74, Ho; 75, Bhumij; 76, Paharia; 77, Kharia; 78, Santal; 79, Oraon; 80, Tadvi; 81, Rathwa; 82, Mixed Caste; 83, Konkan Brahmins; 84, Maratha; 85, Naba-Baudh; 86, Parsi; 87, Katkari; 88, Pawara; 89, Chitpavan Brahmins; 90, Desasth Brahmins; 91, Dhargar; 92, Rajput; 93, Bharia; 94, Maria Gond; 95, Muria; 96, Halba; 97, Saryupari Brahmins; 98, Kamar; 99, Satnami; 100, Kanwar; 101, Nagesia; 102, Havik; 103, Kuruchian; 104, Mullukurunan; 105, Mixed castes (Cochin); 106, Cochin Jews; 107, Toda; 108, Kadar; 109, Irula; 110, Kota; 111, Kurumba; 112, Paniyan; 113, Kadar; 114, Brahmins: Havik; 115, Mukri; 116, Soligas; 117, Koragas; 118, Yerava; 119, Gowda; 120, Kuruva; 121, Christian; 122, Bhovi; 123, Lyngayat; 124, Iyengar; 125, Muslims; 126, Ambalakarar; 127, Iyer; 128, Pallar; 129, Vanniyar; 130, Kadar; 131, Pallar; 132, Vellala; 133, Jennukurumba; 134, Kattunaiken; 135, Mullukurumba; 136, Bettakurumba; 137, Oorali; 138, Baduga; 139, Sakkili; 140, Irula; 141, Kota; 142, Kurumba; 143, Iyengar; 144, Paniyan; 145, Yerukula; 146, Lambadi; 147, Brahmins; 148, Kshatriya; 149, Vysya; 150, Kapu; 151, Madiga; 152, Mala; 153, Relli; 154, Chenchu; 155, Koya; 156, Andh; 157, Pardhi; 158, Thoti; 159, Pardhan; 160, Naikpod Gond; 161, Komati; 162, Reddy; 163, Thogataveera; 164, Yanadi; 165, Yadava; 166, Onge; 167, Great Andamanese; 168, Jarawa; 169, Aka-Bea; 170, Nicobarese.

Thangaraj *et al.* (2006) estimated a coalescent time of around 60,200 ± 8600 yr for M2 and 38,000 ± 10,200 yr for its sub-clade M2a.

The frequency of M2 increases as we progress from Indo-European speaking populations in the north (~6% of M) to Dravidian speaking populations in the south (~13% of M) (Metspalu *et al.* 2004). M2 was reported in a Gowda individual of Karnataka (Rajkumar *et al.* 2005) and two individuals of Thogataveera population from Andhra Pradesh (Sun *et al.* 2006). Although M2 is characteristic of the caste populations of south India, the tribal populations: Chenchu and Koya also show high frequency of M2 (Kivisild *et al.* 2003). A high frequency (19%) of haplogroup M2 was observed among the Austro-Asiatic speaking tribal populations (Basu *et al.* 2003); however, further analysis of the sequences revealed a corrected frequency of about 2% for M2 among the Austro-Asiatic speaking tribal groups of West Bengal (Metspalu *et al.* 2004). This frequency is comparable to M2 frequency (>3%) in Indo-European speaking tribal populations of Punjab and Uttar Pradesh, but significantly lower than its frequency (>14%) among the Dravidian speaking populations of Andhra Pradesh.

M2 was completely absent among the Tibeto-Burman speaking tribals and exhibited high frequency among the tribals of central India (Basu *et al.* 2003). The presence of M2a was reported in individuals from Chenchu population of Andhra Pradesh, Kuruva of Karnataka and Katkari of Maharashtra (Rajkumar *et al.* 2005). The sub-haplogroup also exhibited its presence in Reddy and Thogataveera populations of Andhra Pradesh (Sun *et al.* 2006). M2a has also been reported in Karan and Gope populations of Orissa (Sahoo and Kashyap 2006) and in two individuals of Maratha population from Western India (Gaikwad and Kashyap 2005). Another sub-haplogroup of M2, M2b, defined by the presence of T16357C mutation in addition to the M2 defining basal mutations, was reported in an individual from Kuruva population of Karnataka (Rajkumar *et al.* 2005).

Haplogroup M3

M3 (482-16126) was found to exhibit a high frequency of 9.64% among the Dravidian speaking tribals of south India (Basu *et al.* 2003). In this study, however, the Dravidian speaking tribes were mainly sampled from Tamil Nadu and included Irula, Kota, Kurumba and Toda tribes. M3 haplogroup was also reported among the Chenchu and Koya of Andhra Pradesh (Kivisild *et al.* 2003). It was also the most frequent haplogroup (17%) among the three tribal populations (Pardhan, Naikpod and Andh) (Thanseem *et al.* 2006). Gaikwad and Kashyap (2005) observed M3 in Maratha population of western India, and Sahoo and Kashyap (2006) reported the presence of M3 among the Khandayats of eastern India. M3 was observed even among Paniya from Kerala (Thangaraj *et al.* 2006).

The classification of haplogroup M3, which was previously characterized by a coding region mutation 4580 along

with the control region substitutions 482 and 16,126, was recently revised by Thangaraj *et al.* (2006). Their study suggested that the mutation at 4580 arose later and this was represented as sub-haplogroup M3a. M3a was concentrated in northwestern India (22% among Parsees of Mumbai) suggesting that the region may have been the ancestral source. High frequencies of M3a were also reported among the Rajputs of Rajasthan (14%), Brahmins of Uttar Pradesh (16%) (Metspalu *et al.* 2004), and in Karan and Juang populations of Orissa (Sahoo and Kashyap 2006). The coalescence times of M3 and M3a was calculated to be 27,100 ± 10,200 and 16,400 ± 6100 yr, respectively (Thangaraj *et al.* 2006).

Haplogroup M4

M4 (12007-16311) was estimated to have an overall frequency of ~15% in India with complete absence among Austro-Asiatic speaking tribals and exclusive presence among the Totos of West Bengal (Basu *et al.* 2003). The haplogroup has also been reported among the Dhangars and Chitpavan-Brahmins of Maharashtra of western India (Gaikwad and Kashyap 2005).

Haplogroup M4 shares the 12,007 transition with M18, M30, M37 and M38 and is thus included along with these haplogroups in a super-branch nested in M, named as M4'30. M4 consists of a sub-haplogroup M4a that is defined by two coding (6620 and 7859) and three control region (152, 16145 and 16261) mutations in addition to M4 defining mutation sites (Thangaraj *et al.* 2006). M4a is sparsely spread in most of India with no obvious geographical cline. The sub-haplogroup was found in southeastern Tamil Nadu and Andhra Pradesh but was completely absent among the neighbouring states of Karnataka and Kerala (Metspalu *et al.* 2004). M4a was also observed among caste populations (Oriya Brahmins and Khandayat) and a tribe (Paroja) of Orissa (Sahoo and Kashyap 2006). Thangaraj *et al.* (2006) reported the presence of M4a in Tadwi population of Gujarat. In addition, another sub-haplogroup of M4, M4b (511), was observed in samples of Thogataveera (Andhra Pradesh) and Brahmins (Uttar Pradesh) by Sun *et al.* (2006). The age of M4 haplogroup was estimated to be 25,700 ± 8100 yr and that of its sub-haplogroups, M4a and M4b, was calculated to be 15,400 ± 6300 yr (Thangaraj *et al.* 2006). The coalescence time of M4'30 was computed to be 61,700 ± 12,600 yr (Thangaraj *et al.* 2006).

Haplogroup M5

M5 (1888-16129) was observed to be predominant among the castes rather than tribes (Basu *et al.* 2003). Sun *et al.* (2006) reported haplogroup M5 in individuals from Bhargava and Chaturvedi (Brahmin) populations of Uttar Pradesh and from Thogataveera population of Andhra Pradesh. Individuals from Rajputs of Bihar and Muslims of Karnataka were also found to harbour this haplogroup (Rajkumar *et al.* 2005).

The haplogroup is also present in three predominant caste populations of Maharashtra, namely, Maratha, Desasth Brahmin and Chitpavan Brahmin (Gaikwad and Kashyap 2005) and also among the caste groups of Orissa, viz., Oriya Brahmins, Karan and Gope (Sahoo and Kashyap 2006). But in contrast to the predominance of M5 in caste populations, the tribe Koya displays the presence of M5 upto 32% (Kivisild *et al.* 2003). Even the tribes of Orissa (Juang, Saora and Paroja) show the presence of M5 but in comparatively low frequency of about 5%–10% (Sahoo and Kashyap 2006).

The sub-haplogroup of M5, M5a (709-3921-12477-14323) was reported in individuals from Brahmin populations of Uttar Pradesh and from Reddy populations of Andhra Pradesh (Sun *et al.* 2006). Rajkumar *et al.* (2005) observed this sub-haplogroup in a Bhovi individual from Karnataka of southern India and Thangaraj *et al.* (2006) found M5a in Gadaba population of Orissa. Another sub-haplogroup of M5, M5b was reported in Saora and Paroja populations of Orissa (Sahoo and Kashyap 2006). The coalescence times of M5 and its sub-haplogroup M5a was estimated to be 52,000 ± 14,600 and 23,100 ± 7700 yr, respectively (Thangaraj *et al.* 2006).

Haplogroup M6

M6 (461-3537-5082-5301-5558-9329-10640-13966-14128-16231-16362) is primarily found in the Indus valley and in the western shores of the Bay of Bengal and its sub-haplogroups M6a (16231-16356-16362) and M6b (5585-3486-16188-16231-16362) are concentrated towards the southwest and northeast part of India, respectively (Metspalu *et al.* 2004). M6 was observed in Saora and Paroja tribes of Orissa by Sahoo and Kashyap (2006). The highest frequencies of M6a and M6b were found amongst the Mukri, schedule caste from Karnataka (17%) and in Kashmir (10%), respectively (Metspalu *et al.* 2004). In general, M6a was found to be present in southeastern Tamil Nadu and Andhra Pradesh while absent in neighbouring states, Karnataka and Kerala (Metspalu *et al.* 2004). This was a major subclade among the Chenchu and Koya of Andhra Pradesh too (Kivisild *et al.* 2003). M6a was also reported in Reddy of Andhra Pradesh (Sun *et al.* 2006); Maratha and Dhangar of Maharashtra (Gaikwad and Kashyap 2005) and Brahmins of Orissa (Sahoo and Kashyap 2006). The age of the haplogroup M6 and its sub-haplogroup M6a was estimated to be 27,600 ± 9700 and 23,100 ± 7700 yr, respectively (Thangaraj *et al.* 2006). M6b was reported in Reddy population of Andhra Pradesh by Sun *et al.* (2006) and recently in a sample of Paniya from Kerala (Thangaraj *et al.* 2006). The lineage was found in fairly good frequency among the Pardhan tribe of Andhra Pradesh, in a study by Thanseem *et al.* (2006). The coalescent time for this sub-haplogroup was calculated to be 6000 ± 2100 yr (Metspalu *et al.* 2004).

Haplogroups M18 and M25

The M18 haplogroup was defined by Metspalu *et al.* (2004) based on the transversion at np 16318 and later the definition was revised by Thangaraj *et al.* (2006). Presently, the haplogroup is characterized by two coding region mutations (12498 and 15942) and an additional control region mutation (194). This haplogroup is spread at low frequencies in India, except for the extreme northern region and along the coast of Arabian Sea. High incidence (33%) of the haplogroup was reported in Austro-Asiatic speaking Lodha of West Bengal (Metspalu *et al.* 2004). A relatively high frequency of M18 (13% of M and 8.3% of the total) was observed in Pardhan of Andhra Pradesh, while it was completely absent in Naikpod and Andh tribes (Thanseem *et al.* 2006). M18 was also observed in Brahmins of Uttar Pradesh (Sun *et al.* 2006), Desasth Brahmin of Maharashtra (Gaikwad and Kashyap 2005) and Khandayat of Orissa (Sahoo and Kashyap 2006). Oraon from Bihar was also found to possess this haplogroup (Thangaraj *et al.* 2006). The coalescence age of this haplogroup was estimated to be 20,800 ± 8900 yr (Thangaraj *et al.* 2006).

Haplogroup M25 was tentatively named by Metspalu *et al.* (2004) based on a transition at np 16304 combined with G15928A mutation. Compared to M18, M25 is moderately frequent in Kerala and Maharashtra but hardly found in other parts of the subcontinent. The haplogroup was reported in an individual of Naikpod Gond of Andhra Pradesh (Rajkumar *et al.* 2005); Dhangar and Desasth Brahmins of Maharashtra (Gaikwad and Kashyap 2005) and Khandayat of Orissa (Sahoo and Kashyap 2006). The lineage constituted 9.5% of the total M lineages in Naikpods (Thanseem *et al.* 2006). Coalescence estimates for this haplogroup is about 20,100 ± 6800 yr (Thangaraj *et al.* 2006).

Haplogroup M30

Rajkumar *et al.* (2005) identified and designated a new lineage M30 (G12007A) and M30a (T195A and G15431A) based on complete mitochondrial genome sequencing of 23 Indian samples. The mutations characterizing this lineage was observed in five samples from eastern part of India, viz., Bihar (Kurmi, Yadav and Baniya), West Bengal (Mahishya) and Orissa (Saora) and two samples from south India (Christians of Karnataka and Lambadi of Andhra Pradesh). However, later the definition of M30 was narrowed down by Sun *et al.* (2006) and M30 was later identified by the mutations T195A and G15431A besides G12007A. The study recognized four sub-haplogroups of M30 viz, M30a, M30b, M30c, M30d based on the observed mutations sites.

An individual from Reddy population of Andhra Pradesh was categorized into M30a; a sample of Thogataveera from Andhra Pradesh was classified into M30b; M30c was found in Thogataveera of Andhra Pradesh and Chaturvedi of Uttar Pradesh; M30d was identified among Bhargava of Uttar Pradesh and Thogataveera of Andhra Pradesh. Sahoo and

Kashyap (2006) reported haplogroup M30 in castes (Oriya Brahmin, Karan, Khandayat and Gope) and tribes (Juang and Saora) of Orissa. Saora exhibited a high frequency of ~32% of M30 followed by Karan (~24%), Oriya Brahmin (~20%) and Juang (~20%), while Khandayat and Gope showed a lower frequency (~6%) of M30. M30 was also present at low frequencies (1.5%–2.5%) in Pardhan, Naikpod Gond and Andh tribal populations of Andhra Pradesh (Thanseem *et al.* 2006). The coalescence time of haplogroup M30 was estimated to be 15,400 ± 6300 yr and that of its sub-haplogroups M30a and M30c were calculated to be 5100 ± 3600 yr. The ages of M30b and M30d were computed to be 4177 ± 2800 and 12,800 ± 5700 yr, respectively (Thangaraj *et al.* 2006).

Haplogroups M31 and M32

Rajkumar *et al.* (2005) defined another new haplogroup, M31, characterized by a substitution at A5319G, in four samples, of Karnataka (Iyengar Brahmin and Lingayat), Andhra Pradesh (Komati) and Maharashtra (Pawar). Thangaraj *et al.* (2005a) designated two new clades M31 (4907-11176-15440-15530) and M32 (3817-9064-12189-15754) by the analysis of the complete mtDNA sequence of Onge and Great Andamanese and suggested an *in situ* origin of these haplogroups on the Andaman Island. The coalescence time of M31 and M32 was estimated to be 3000 ± 2000 ybp and 12000 ± 4000 ybp, respectively (Thangaraj *et al.* 2005a).

The Andaman-specific lineage M31 suggests two clear-cut population-specific subclades. Onge and Jarwa share M31a1 branch (200 and 13710), while M31a2 (9617) clade is found only among the Great Andamanese individuals (Thangaraj *et al.* 2006). Based on the comments of Palanichamy *et al.* (2006), Thangaraj *et al.* (2006) reconstructed the classification of M31 lineage by including a Rajbanshi sequence from Palanichamy *et al.* (2006). The Rajbanshi sample was categorized into the sub-haplogroup M31b (152-808-3337-3834-8092-14212-15440-15667-16126-16136) based on the coding region substitutions.

Later, Endicott *et al.* (2006) suggested some alterations in the phylogeny of M31 (4907-11176-15440-15530-16126) and M32 (143-195-207-3817-9064-12189-16319-16344-16357) haplogroups, based on the discovery of some novel mutations and also revised the definition of the Andaman-specific M31a1 and M32 lineages. The phylogeny of haplogroup M31 now consists of M31a (3999-12876) and M31b (808-3337-3834-8092-14212-15876-16136), of which M31b was reported in a Rajbanshi individual. M31a in turn includes M31a1 (249del-1524-2045-3975-8973-9581-11014-14407-16126-16311) and M31a2 (195-16093-16145). Two sub-clades of M31a1 were defined viz., M31a1a (9617) and M31a1b (13710) among the speakers of Great Andamanese and Onge-Jarawa dialects, respectively. M31a1b1 of sub-clade M31a1b, was defined in a Jarawa individual based on the coding region mutation 8108 apart from the mutations characterizing M31a1 and M31a1b lin-

eages. M31a2 was reported among the Lodha, Chenchu and Lambadi individuals. Recently, Reddy *et al.* (2007) reported a novel sub-haplogroup M31c (188-234-282-9269-14152-15300-15935-16136-16311), under the haplogroup M31, among the Austro-Asiatic Khasi populations (~5%), with a maximum frequency among the Bhoi (~17%). The study also suggested a TMRCA of ~40,000 ybp for M31 haplogroup. Haplogroup M32 includes M32a (15754) and its sub-clade M32a1 (8108), reported among Great Andamanese and Onge-Jarawa speakers, respectively.

Haplogroup M33

Haplogroup M33 is defined by the coding region substitution at np 2361. Sun *et al.* (2006) reported this haplogroup in Rajbansi population of West Bengal and Brahmins of Uttar Pradesh. The haplogroup was reported in Naikpod and Andh populations (~3 to 3.4%) but was completely absent in Pardhan (Thanseem *et al.* 2006). A sub-haplogroup of M33, M33a (8562-15908), was reported in Tadvi population of Gujarat (Thangaraj *et al.* 2006) and among Khasi-Khumic populations (~5%), with an exceptionally high frequency (~55%) among the Garo of Meghalaya (Reddy *et al.* 2007). Recently, a novel sub-haplogroup M33b (1719-3221-6293-16324) was reported by Reddy *et al.* (2007) among Pnar (~22%), Lyngngam (~2%), Khyriam (~3%) and Garo (~3%) tribes. The TMRCA of haplogroup M33 was reported to be ~50,000 ybp (Reddy *et al.* 2007).

Haplogroups M34, M35, M36, M37, M38, M39 and M40

Sun *et al.* (2006) observed the motif 569-3010-6794-11101-15865-16249 in two mtDNAs of Brahmins of Uttar Pradesh, and categorized the motif into haplogroup M34. Thangaraj *et al.* (2006) reported sub-haplogroup M34a, in a random sample drawn from Karnataka, characterized by six coding region (3447-8404-10361-11992-12311-14094) and three control region mutations (at sites 146, 16095 and 16359). The coalescence time of the haplogroups M34 and M34a were computed to be around 28,400 ± 8300 and 15,400 ± 6300 yr, respectively (Thangaraj *et al.* 2006).

Haplogroup M35 is recognized by two mutations (at sites 199 and 12561) and its sub-haplogroup M35a is characterized by another five mutations (at sites 482, 5432, 10670, 15924 and 16093). One sample from Thogataveera and two samples from Reddy of Andhra Pradesh were categorized into haplogroups M35 and M35a, respectively (Sun *et al.* 2006). M35 was also reported in Pardhan (0.52%) and Andh (6.06%) tribal populations of Andhra Pradesh (Thanseem *et al.* 2006). The coalescence times of the haplogroups M35 and M35a were estimated to be 25,700 ± 8100 and 10,300 ± 5100 yr, respectively (Thangaraj *et al.* 2006).

Three specific mutations (at sites 239, 7271 and 15110) define haplogroup M36 and this was reported in two samples of Thogataveera population from Andhra Pradesh (Sun *et al.* 2006). The coalescence time of the haplogroup was calculated to be around 33,400 ± 9300 yr (Thangaraj *et al.* 2006).

Samples from Brahmins of Uttar Pradesh and Reddy of Andhra Pradesh shared mutations 12007 and 10556, and hence were categorized into a novel haplogroup M37 (Sun *et al.* 2006). M37a, a sub-haplogroup of M37 was reported in a sample of Rathwa (Gujarat) based on a single coding (7853) and two control region (151 and 152) mutations (Thangaraj *et al.* 2006). The ages of the haplogroup M37 and sub-haplogroup M37a were estimated to be 28,300 ± 11,900 and 18,000 ± 6800 yr, respectively (Thangaraj *et al.* 2006).

The motif of nine mutations, at sites 189, 246, 1808, 6367, 6899, 9966, 12007, 15314 and 15497, designates haplogroup M38 (Sun *et al.* 2006). The haplogroup was observed in Bhargava of Uttar Pradesh and Thogataveera of Andhra Pradesh (Sun *et al.* 2006). Recently, Thanseem *et al.* (2006) reported this haplogroup in Pardhan (2.07%) and Naikpod Gond (1.14%) populations of Andhra Pradesh. The coalescence time of M38 was computed to be 21,000 ± 7100 yr (Thangaraj *et al.* 2006).

Haplogroup M39, characterized by a pronounced HVS-II motif (55+T, 65+T and 66T) and three coding region mutations (at sites 1811, 8679 and 15938), was observed in Bhargava of Uttar Pradesh, Reddy and Thogataveera of Andhra Pradesh (Sun *et al.* 2006). Thanseem *et al.* (2006) observed a frequency of 12.12% of M39 in Andh tribal population of Andhra Pradesh, but the haplogroup was completely absent among the other two studied populations, Pardhan and Naikpod Gond. The coalescence time of M39 was calculated to be 27,600 ± 7300 yr (Thangaraj *et al.* 2006).

Haplogroup M40 consists of a specific motif composed of four mutations (at sites 8925, 15721, 15954 and 16463) and was observed in the samples of Reddy and Thogataveera populations of Andhra Pradesh. M40a, a sub-haplogroup of M40, was defined by Thangaraj *et al.* (2006). The sub-haplogroup is defined by a coding (13542) and three control region (200-16179-16294) mutations. Thanseem *et al.* (2006) reported the presence of M40 in Pardhan (0.52%) of Andhra Pradesh, however the haplogroup was absent in Naikpod Gond and Andh populations that were included in the study. The ages of the haplogroup M40 and sub-haplogroup M40a were computed to be 24,100 ± 7700 and 18,000 ± 6800 yr, respectively (Thangaraj *et al.* 2006).

Haplogroup M41

Haplogroup M41 (six coding, 870-6297-12398-12469-13656-15601 and three control region, 375-16327-16330 mutations) was reported in the samples of Pardhan from Andhra Pradesh (Thangaraj *et al.* 2006). Recently, Thanseem *et al.* (2006) also reported the presence of M41 in Pardhan of Andhra Pradesh (3.11%). The other two studied populations, Andh and Naikpod Gond, showed complete absence of this haplogroup. The haplogroup was tentatively classified into three sub-haplogroups M41a, M41b and M41c based on the mutation sites observed. The age of M41 was calculated to be 12,800 ± 5700 yr (Thangaraj *et al.* 2006).

Haplogroups M48, M49 and M50

Three novel haplogroups, M48, M49 and M50, were recently defined by Reddy *et al.* (2007) among the Austro-Asiatic Khasi tribes. Haplogroup M48 (199-6366-15900-16225-16234-16390) was found at a frequency of about 11% among the Austro-Asiatic Khasi groups, ranging from zero in War-Jaintia to as high as 26% in Lyngngam, with a frequency of 4% among the Garo of Meghalaya. Haplogroups M49 (7598-11263-12346-14384-15511-15647-16153-16234) and M50 (1383-3417-6092-7226-8802-12281-12486-13940-15300-15055-15660-16166C-16263-16309) were found at an average frequency of about 3% in the Khasi populations and was completely absent in the Garo as well as in some of the subgroups of Khasi.

Other minor haplogroups of M

Other M sub-haplogroups that are common in mainland east Asia (M7, M8 (including C, Z), M9, D, G) have also been reported in India at relatively lower frequencies. Haplogroup M8 is characterized by the transitions at 4715, 8584, 16298 and transversions at 7196 and 15487 and consists of sub-haplogroups C, Z and M8a (Yao *et al.* 2002). M8c has been reported in samples from Assam (2), Himachal Pradesh (1), Karnataka (1), Kerala (6), Nagaland (5), Tamil Nadu (2), Uttar Pradesh (1) and Tripura (6) and M8z have been reported in sample from Assam (1), Himachal Pradesh (1) and Tripura (3); whereas M8a have not yet been reported (Metspalu *et al.* 2004). M9 is defined by a transition at 4491 (Yao *et al.* 2002) and consists of two sub-haplogroups M9E and M9a. The sub-haplogroup M9a have been reported in samples from Arunachal Pradesh (2), Assam (3), Himachal Pradesh (1), Tripura (3) and Uttar Pradesh (1); whereas M9E was completely absent (Metspalu *et al.* 2004). M10 haplogroup, first defined by Yao *et al.* (2002) in Han Chinese, was reported by Metspalu *et al.* (2004) in a single individual from Assam. Haplogroup D (16362, 5178A) was categorized into sub-haplogroups D4 and D5 which was further classified into D4a, D4b and D5a, respectively by Yao *et al.* (2002). Haplogroup D was reported in samples of Andhra Pradesh (1), Arunachal Pradesh (4), Assam (6), Himachal Pradesh (2), Nagaland (4), Punjab (1), Tripura (17) and Uttar Pradesh (4) of which sub-haplogroup D4 was present in samples from Assam (2), Himachal Pradesh (1), Nagaland (3), Tripura (3) and Uttar Pradesh (2). The D4 sample from Himachal Pradesh was further categorized into D4b based on the coding region mutations. D5a, sub-haplogroup of D5 was observed in Assam (2), Arunachal Pradesh (2) and Tripura (7) of northeastern part of India. Haplogroup G was reported in Assam (2), Himachal Pradesh (5), Kerala (1), Uttar Pradesh (2) and Tripura (1). Haplogroup G was further categorized into their respective sub-haplogroups; G2 was present in Assam, Himachal Pradesh and Uttar Pradesh, and G3 was reported in Himachal Pradesh (Metspalu *et al.* 2004).

Overview

In recent years, many molecular genetic evidences from several populations inhabiting the Indian subcontinent has been obtained to understand the unresolved population genetic issues related to the Indian populations and the peopling of the subcontinent (Mountain *et al.* 1995; Kivisild *et al.* 1999, 2003; Roychoudhury *et al.* 2000, 2001; Basu *et al.* 2003; Cordaux *et al.* 2003, 2004; Thangaraj *et al.* 2003, 2005a,b, 2006; Palanichamy *et al.* 2004; Sun *et al.* 2006). Though the number of studies (~ 170 populations) contributed are scanty, these studies do provide us glimpse of some clear trends of the mtDNA diversity and help in investigating further insights into the past genetic history of man in India.

A majority of the populations (~ 60%) belong to macrohaplogroup M, and a higher frequency of this macrohaplotype is observed among tribes as compared to caste populations. Several India-specific haplogroups (~ 22) of M have been found, and are probably of *in situ* origin. The most ancient haplogroup M2 is reported among the Austro-Asiatic speaking Mundari, Dravidian (southern), and Indo-European speaking castes and tribes, but is completely absent among Tibeto-Burman speakers. M3 is predominant among southern Dravidian tribes; M4 is absent in Austro-Asiatic tribes and M5 is more a caste-specific haplogroup. Haplogroups M31 and M32 are mainly found among Andaman tribes, but however, the sub-haplogroup of M31, M31b shows presence in a Rajbanshi (West Bengal) sample and another sub-haplogroup M31c is present among the Khasi-Khmuc speaking Austro-Asiatic tribes. Some novel haplogroups especially M48, M49, M50 are found among Khasi sub-tribes of the northeastern hills.

The overall distribution of sub-haplogroups of M among castes and tribes is shown in figure 3. A majority of mtDNA M sub-haplogroups pervades among the castes and tribes, though some subhaplogroups differ between the two, and some are more specific to either castes or tribes. For example, a few haplogroups such as M25, M35, M38 and M39 show almost similar frequencies among castes and tribes, although these are found at low frequencies (less than 5%). The haplogroups M34 and M36 are observed only among castes, whereas six haplogroups: M32, M34, M41, M48, M49 and M50 are found only in tribes. While some haplogroups are more predominant in castes and some others are prevalent among tribes. For example, five haplogroups: M2, M3, M18, M31 and M33 show higher frequencies in tribes than castes and five other haplogroups: M4, M5, M6, M30 and M37 are more prevalent among castes than in tribes. Of the observed haplogroups, M5 and M31 showed significant differences (Z-test, at 95% confidence interval) among castes and tribes. Figure 4 shows the region-wise and language-wise distribution of the M haplogroups among the studied castes and tribes of India. Interestingly, the two branches of the Austro-Asiatic linguistic family exhibit differences in their haplogroup diversity: the Mon-Khmer branch of north-eastern region showing greater haplogroup diversity than the Mundari branch of the eastern region. Also, the tribes of the same linguistic family distributed over different geographical regions show differences in their haplogroup diversity. Dravidian tribes of the southern region were found to show more number of haplogroups of M than those of the central and the eastern regions. The Andaman and Nicobar tribes show unique haplogroup composition, different from

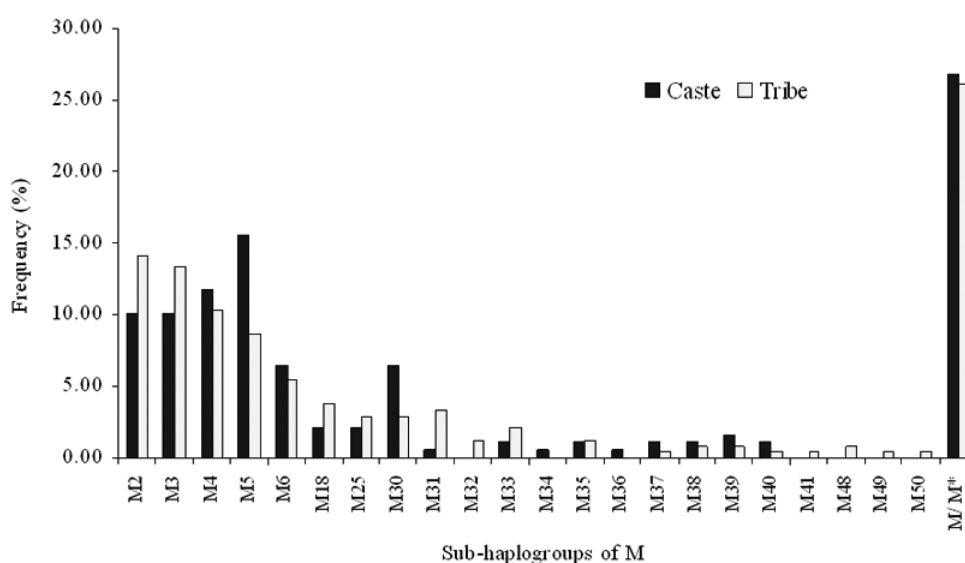


Figure 3. Frequency distribution of the sub-haplogroups of M among castes and tribes of India.

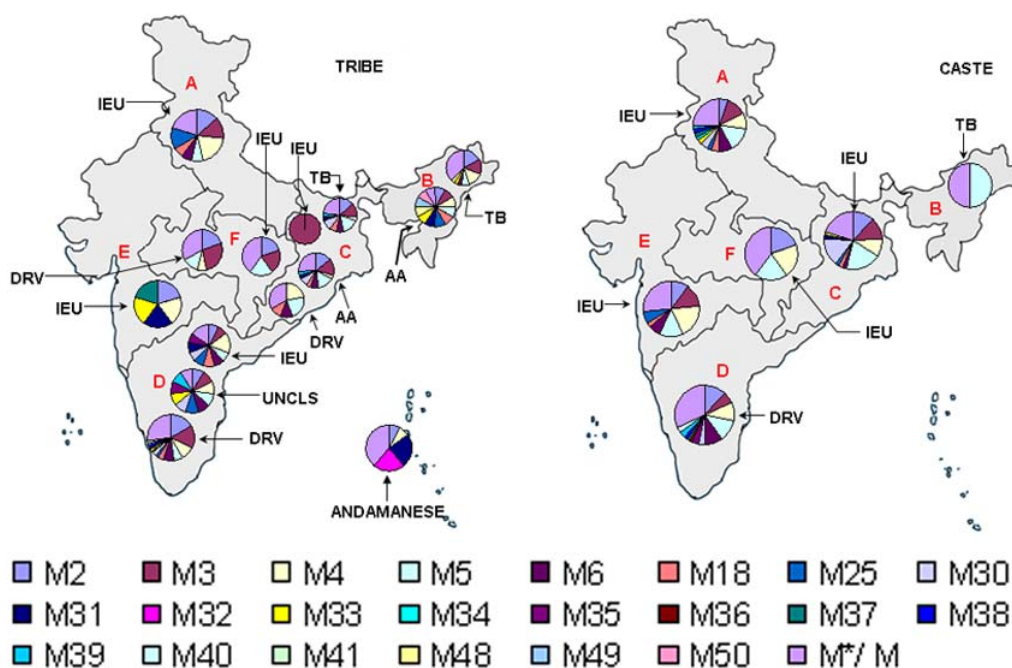


Figure 4. Distribution of the sub-haplogroups of M across the six regions of India. A, Northern; B, North-Eastern; C, Eastern; D, Southern; E, Western; F, Central IEU, Indo-European (IEU); TB, Tibeto-Burman; AA, Austro-Asiatic; DRV, Dravidian

the other tribal populations of the subcontinent. Among the castes, the Indo-European speakers of the northern region and the Dravidian speakers of the southern region show more haplogroup diversity than the caste populations of other regions.

The Indian picture of M haplogroup distribution shows wide prevalence of some haplogroups across the regions and ethnicity, though some show affiliation to regions and ethnicity. The uneven distribution of M sub-haplogroups across linguistic, ethnic and geographic boundaries could possibly be explained by the (i) low molecular resolution: since few general markers are being used to predate the population subdivisions, every population becomes similar and also, if the population diversification is relatively recent, there is not enough time to generate and accumulate mutations; (ii) high level of interbreeding and admixture among populations; (iii) substructuring of populations into small endogamous units, each of which evolves independently and the genetic drift erases the general genetic signatures of population histories. Possibly, a better picture can be expected to emerge in the future with the addition of more representative and adequate samples of the total of 4635 communities.

It is well established that mtDNA data helps in understanding the past genetic history of man and especially the maternal migration of the populations that shaped the global human populations. Among the M haplogroups, M2 is the most ancient (the estimated coalescence time around $60,200 \pm 8600$ yr) and is found mostly among the Dra-

vidian castes (with the exception of their presence among Chenchu and Koya tribes) and Mundari speaking Austro-Asiatic tribes. This estimated coalescence time of M2 corroborates with the 'Out of Africa' theory of human evolution. The complex mtDNA tree could be the result of population expansion. The absence of M2 and the presence of east/southeast Asian specific mtDNA haplogroups (of macrohaplogroup N) among the Tibeto-Burman speakers of northeast India suggest their independent origin and relatively recent entry possibly from east/southeast Asian regions.

There are several limitations in the studies that might influence the inferences drawn. The implemented sampling strategy is one of the major constraints, as the selected samples do not always wholly represent the studied populations with reference to their geographical settlement as well as their size. Further, constraint also prevails in the laboratory techniques and the lack of quality control in sequencing and identification of mutation sites. For example, many published sequences on scrutiny were found to be erroneous (Sun *et al.* 2006 on the study of Rajkumar *et al.* 2005). The study based on about 3.6% of the Indian populations so far reveal vast mtDNA diversity and several unique haplogroups as a result of which the mtDNA phylogenetic tree depicts a very complex scenario that lead to difficulty in discerning any clear patterns. With increasing studies involving more populations, the mtDNA phylogenetic tree for the Indian subcontinent may be expected to become even more complex.

This raises the necessity for developing improved methods of classification of the mtDNA haplogroups to obtain a better clarity of the mtDNA diversity of the Indian populations.

References

- Anderson S., Bankier A. T., Barrell B. G., de Bruijn M. H., Coulson A. R., Drouin J. et al. 1981 Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.
- Andrews R. M., Kubacka I., Chinnery P. F., Lightowlers R. N., Turnbull D. M. and Howell N. 1999 Reanalysis and revisions of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147.
- Ballinger S. W., Schurr T. G., Torroni A., Gan Y. Y., Hodge J. A., Hassan K. et al. 1992 Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics* **130**, 139–152.
- Bamshad M. J., Watkins W. S., Dixon M. E., Jorde L. B., Rao B. B., Naidu J. M. et al. 1998 Female gene flow stratifies Hindu castes. *Nature* **395**, 651–652.
- Bamshad M., Kivisild T., Watkins W. S., Dixon M. E., Ricker C. E., Rao B. B. et al. 2001 Genetic evidence on the origins of Indian caste populations. *Genome Res.* **11**, 994–1004.
- Basu A., Mukherjee N., Roy S., Sengupta S., Banerjee S., Chakraborty M. et al. 2003 Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res.* **13**, 2277–2290.
- Brown W. M. 1980 Polymorphism in mitochondrial DNA of human as revealed by restriction endonuclease analysis. *Proc. Natl. Acad. Sci. USA* **77**, 3605–3609.
- Cann R. L., Stoneking M. and Wilson A. C. 1987 Mitochondrial DNA and human evolution. *Nature* **325**, 31–36.
- Chaubey G., Metspalu M., Kivisild T. and Villems R. 2007 Peopling of South Asia: investigating the caste-tribe continuum in India. *BioEssays* **29**, 91–100.
- Chen Y. S., Torroni A., Excoffier L., Santachiara-Benerecetti A. S. and Wallace D. C. 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 G. R., Aunger R., Sirajuddin S. M. and Stoneking M. 2003 Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur. J. Hum. Genet.* **11**, 253–264.
- Cordaux R., Weiss G., Saha N. and Stoneking M. 2004 The north-east Indian passageway: a barrier or corridor for human migration? *Mol. Biol. Evol.* **21**, 1525–1533.
- Denaro M., Blanc H., Johnson M. J., Chen K. H., Wilmsen F., Cavalli-Sforza L. L. and Wallace D. C. 1981 Ethnic variation in *HpaI* endonuclease cleavage patterns of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **78**, 5768–5772.
- Elson J. L., Andrews R. M., Chinnery P. F., Lightowlers R. N., Turnbull D. M. and Howell N. 2001 Analysis of European mtDNAs for recombination. *Am. J. Hum. Genet.* **68**, 145–153.
- Endicott P., Gilbert T. P., Stringer C., Laiueza-Fox C., Willerslev E., Hansen J. A. and Cooper A. 2003 The genetic origins of the Andaman Islanders. *Am. J. Hum. Genet.* **72**, 178–184.
- Endicott P., Metspalu M., Stringer C., Macaulay V., Cooper A. and Sanchez J. J. 2006 Multiplexed SNP typing of ancient DNA clarifies the origin of Andaman haplogroups amongst south Asian tribal populations. *PLoS One* **1**, e81.
- Excoffier L. and Langaney A. 1989 Origin and differentiation of human mitochondrial DNA. *Am. J. Hum. Genet.* **44**, 73–85.
- Finnilä S. and Majamaa K. 2001 Phylogenetic analysis of mtDNA haplogroup TJ in a Finnish population. *J. Hum. Genet.* **46**, 64–69.
- Finnilä S., Hassinen I. E., Ala-Kokko L. and Majamaa K. 2000 Phylogenetic network of the mtDNA haplogroup U in northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis. *Am. J. Hum. Genet.* **66**, 1017–1026.
- Finnilä S., Lehtonen M. S. and Majamaa K. 2001 Phylogenetic network for European mtDNA. *Am. J. Hum. Genet.* **68**, 1475–1484.
- Gaikwad S. and Kashyap V. K. 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.
- Hammer M. F., Karafet T., Rasanayagam A., Wood E. T., Altheide T. K., Jenkins T. et al. 1998 Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* **15**, 427–441.
- Hebsgaard M. B., Wiuf C. M., Gilbert T. P., Glenner H. and Willerslev E. 2007 Evaluating Neanderthal genetics and phylogeny. *J. Mol. Evol.* **64**, 50–60.
- Hedges S. B., Kumar S., Tamura K. and Stoneking M. 1992 Human origins and analysis of mitochondrial DNA sequences. *Science* **255**, 737–739.
- Ingman M. and Gyllensten U. 2001 Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. *J. Hered.* **92**, 454–461.
- Ingman M., Kaessmann H., Pääbo S. and Gyllensten U. 2000 Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713.
- Johnson M. J., Wallace D. C., Ferris S. D., Ratazzi M. C. and Cavalli-Sforza L. L. 1983 Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. *J. Mol. Evol.* **19**, 255–271.
- Kivisild T., Bamshad M. J., Kaldma K., Metspalu M., Metspalu E., Reidla M. et al. 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. et al. 2003 The genetic heritage of earliest settlers persist in both the Indian tribal and caste populations. *Am. J. Hum. Genet.* **72**, 313–332.
- Kong Q-P., Yao Y-G., Sun C., Bandelt H-J., Zhu C-L. and Zhang Y-P. 2003 Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am. J. Hum. Genet.* **73**, 671–676.
- Kumar V., Langsith B. T., Biswas S., Babau J. P., Rao T. N., Thangaraj K. et al. 2006 Asian and non-Asian origins of Mon-Khmer and Mundari speaking Austro-Asiatic populations of India. *Am. J. Hum. Biol.* **18**, 461–469.
- Maca-Meyer N., González A. M., Larruga J. M., Flores C. and Cabrera V. C. 2001 Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.* **2**, 13.
- Maca-Meyer N., González A. M., Pestano J., Flores C., Larruga J. M. and Cabrera V. C. 2003 Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. *BMC Genet.* **4**, 15.
- Macaulay V., Richards M., Hickey E., Vega E., Cruciani F., Guida V. et al. 1999 The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **64**, 232–249.
- Macaulay V., Hill C., Achilli A., Rengo C., Clarke D., Meehan W. et al. 2005 Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308**, 1034–1036.
- Maddison D. R., Ruvolo M. and Swofford D. L. 1992 Geographic origins of human mitochondrial DNA: phylogenetic evidence from control region sequences. *Syst. Biol.* **41**, 111–124.

- Metspalu M., Kivisild T., Metspalu E., Parik J., Hudjashov G., Kaldma K. *et al.* 2004 Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet.* **5**, 26.
- Mishmar D., Ruiz-Pesini E., Golik P., Macaulay V., Clark A.G., Hosseini S. *et al.* 2003 Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci.* **100**, 171–176.
- Mountain J. L., Hebert J. M., Bhattacharyya S., Underhill P. A., Ottolenghi C., Gadjil M. and Cavalli-Sforza L. L. 1995 Demographic history of India and mtDNA-sequence diversity. *Am. J. Hum. Genet.* **56**, 979–992.
- Palanichamy M. G., Sun C., Agrawal S., Bandelt H. J., Kong Q. P. and Khan F. 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.
- Palanichamy M. G., Agrawal S., Yao Y. G., Kong Q. P., Sun C., Khan F. *et al.* 2006 comment on “Reconstructing the origin of Andaman islanders”. *Science* **311**, 470.
- Quintana-Murci L., Semino O., Bandelt H.-J., Passarino G., McElreavey K. and Santachiara-Benerecetti A. S. 1999 Genetic evidence of an early exit of *Homo sapiens* from Africa through eastern Africa. *Nat. Genet.* **23**, 437–441.
- Quintana-Murci L., Chaix R., Wells R. S., Behar D. M., Sayar H. and Scozzari R. 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 H. B., Trivedi R. and Kashyap V. K. 2005 Phylogeny and antiquity of M macrohaplogroup inferred from complete mtDNA sequence of Indian specific lineages. *BMC Evol. Biol.* **5**, 26.
- Ratnagar S. 1995 Archaeological perspectives of early Indian societies. In *Recent perspectives of early Indian history* (ed. R. Thapar), pp. 1–52. Popular Prakashan, Mumbai.
- Rogers A. R. and Jorde L. B. 1995 Genetic evidence on modern human origins. *Hum. Biol.* **67**, 1–36.
- Reddy B. M., Langsteih B. T., Kumar V., Nagaraja T., Reddy A. N. S. and Meka A. 2007 Austro-Asiatic tribes of northeast India provide hitherto missing genetic link between south and southeast Asia. *PLoS One* **2**, e1141.
- Roychoudhury S., Roychoudhury S., Roy S., Dey B., Chakraborty M., Roy M. *et al.* 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 M. V. *et al.* 2001 Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum. Genet.* **109**, 339–350.
- Sahoo S. and Kashyap V. K. 2006 Phylogeography of Mitochondrial DNA and Y-Chromosome haplogroups reveal asymmetric gene flow in populations of Eastern India. *Am. J. Phys. Anthropol.* **131**, 84–97.
- Salas A., Richaqrds M., De la Fe T., Lareu M. V., Sobrino B., Sanchez-Diz P. *et al.* 2002 The making of the African mtDNA landscape. *Am. J. Hum. Genet.* **71**, 1082–1111.
- Schurr T. G., Ballinger S. W., Can Y. Y., Hodce J. A., Merriwether D. A., Lawrence D. N. *et al.* 1990 Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am. J. Hum. Genet.* **46**, 613–623.
- Scozzari R., Torroni A., Semino A., Sirugo G., Brega A. and Santachiara-Benerecetti A. S. 1988 Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms. *Am. J. Hum. Genet.* **43**, 534–544.
- Stringer C. B. and Andrews P. 1988 Genetic and fossil evidence for the origin of modern humans. *Science* **239**, 1263–1268.
- Sun C., Kong Q. P., Palanichamy M. G., Agrawal S., Bandelt H. J., Yao Y. G. *et al.* 2006 The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as referred from complete genomes. *Mol. Biol. Evol.* **23**, 683–690.
- Templeton A. R. 1992 Human origins and analysis of mitochondrial DNA sequences. *Science* **255**, 737.
- Templeton A. 2002 Out of Africa again and again. *Nature* **416**, 45–51.
- Thangaraj K., Singh L., Reddy A. J., Rao R., Sehgal C. S., Underhill P. A. *et al.* 2003 Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr. Biol.* **13**, 86–93.
- Thangaraj K., Chaubey G., Kivisild T., Reddy A. G., Singh V. K., Rasalkar A. and Singh L. 2005a Reconstructing the origin of Andaman Islanders. *Science* **308**, 996.
- Thangaraj K., Sridhar V., Kivisild T., Reddy A. G., Chaubey G. and Singh V. K. 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.
- Thangaraj K., Chaubey G., Singh V., Vanniarajan A., Thanseem I., Reddy A. G. and Singh L. 2006 In situ origin of deep rooting lineages of mitochondrial macrohaplogroup M in India. *BMC Genomics* **7**, 151.
- Thanseem I., Thangaraj K., Chaubey G., Singh V. K., Bhaskar V. K. S. L. and Reddy B. M. 2006 Genetic affinities among the lower caste and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC Genet.* **7**, 42.
- Thapar R. 1995 The first millennium B. C. in northern India (up to the end of Mauryan period). In *Recent perspectives of early Indian history* (ed. R. Thapar), pp. 80–141. Popular Prakashan, Mumbai.
- Torroni A., Huoponen K., Francalacci P., Petrozzi M., Morelli L. and Scozzari R. 1996 Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**, 1835–1850.
- Torroni A., Bandelt H. J., D’Urbano L., Lahermo P., Moral P., Sellitto D. *et al.* 1998 mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am. J. Hum. Genet.* **62**, 1137–1152.
- Torroni A., Rengo C., Guida V., Cruciani F., Sellitto D., Coppa A. *et al.* 2001. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am. J. Hum. Genet.* **69**, 1348–1356.
- Vigilant L., Stoneking M., Harpending H., Hawkes K. and Wilson A. C. 1991 African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503–1507.
- Watkins W. S., Bamshad M., Dixon M. E., Bhaskara Rao B., Naidu J. M., Reddy P. G. *et al.* 1999 Multiple origin of the mtDNA 9 bp deletion in populations of South India. *Am. J. Phys. Anthropol.* **109**, 147–158.
- Watson E., Forster P., Richards M. and Bandelt H. J. 1997 Mitochondrial footprints of human expansions in Africa. *Am. J. Hum. Genet.* **61**, 691–704.
- Yao Y. G., Kong Q. P., Bandelt H. J., Kivisild T. and Zheng Y. P. 2002 Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am. J. Hum. Genet.* **70**, 635–651.

Received 3 April 2008, in revised form 13 June 2008; accepted 28 July 2008

Published on the Web: 13 March 2009