

## Molecular Genetic Perspective of Indian Populations: A Y-Chromosome Scenario

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

### INTRODUCTION

The intriguing aspect of the Indian populations is the immense cultural and biological diversity of the various castes, tribes and other ethnic groups (e.g. religious communities). Comprehending the possible reasons for this extensive diversity and tracing the history of the populations, viz., the origin, migration and settlement, has been the focus of study for archeologists, historians, linguists, anthropologists and of late for molecular geneticists. In this regard, various competing hypotheses have been put forward to explain the evolutionary history of the Indian tribes as well as castes.

In case of the tribal populations, the proto-Australoid tribes belonging to the Austric linguistic groups are believed to be the basic element in the Indian population (Thapar, 1966). Many other anthropologists, historians and linguists (Pattanayak, 1998; Rapson, 1955; Rishley, 1915) have also supported the view that the Austro-Asiatic (a subfamily of the Austric language family) speaking tribes to be the original inhabitants of India. Some other scholars (Buxton, 1925; Sarkar, 1958) have proposed that the Dravidians are the original inhabitants, and that the Austro-Asiatics are later immigrants. The Tibeto-Burman speaking tribals, primarily inhabiting northeastern regions of India, are thought to be the latest immigrants among the tribals to India from Tibet and Myanmar (Guha, 1935).

With regard to the caste populations, linguists of the 19<sup>th</sup> and 20<sup>th</sup> centuries proposed that the extant castes are descendants of primarily West Eurasians who migrated from Northwest about 3000–8000 years ago (Poliakov, 1974; Renfrew, 1989). These nomadic migrants may have purportedly admixed with or displaced the indigenous Dravidic-speaking proto-Asian populations and subsequently established the Hindu caste hierarchy (Poliakov, 1974; Cavalli-Sforza et al., 1994) and placed themselves in predominantly castes of higher rank. However, this was hardly supported by the archaeological evidence (Shaffer, 1982).

Anthropologists also attempted to understand the structure of Indian populations and its vast diversity, by employing certain anthropological variables as well as classical genetic markers. Dendrograms constructed on the basis of these biological traits depicted clear differentiation of Indian populations, based on different geographical regions, ethnicity and linguistic families. Tribal populations differed significantly from castes, and within castes the lower ranks behaved differently from that of the higher ranks. Among tribal populations, the southern groups were quite different from the central and northern groups. Geographic proximity was found to play a key role in determining the genetic affinities among the Indian populations (Bhasin and Walter, 2001; Malhotra and Vasulu, 1993). However, another study by Majumder and Mukherjee (1993) on Indian populations based on classical genetic traits yielded contrasting results; genetic variation was not found to correlate well with the social rank among the caste populations. Also, geographical proximity was not found to always correlate well with genetic affinity among the populations. There was no clear pattern of genetic variation and affinities in Indian populations that can be assigned to any particular factor. These contradictory inferences drawn about Indian populations from different studies based on classical genetic traits can possibly be addressed and substantiated by using high-resolution molecular genetic markers.

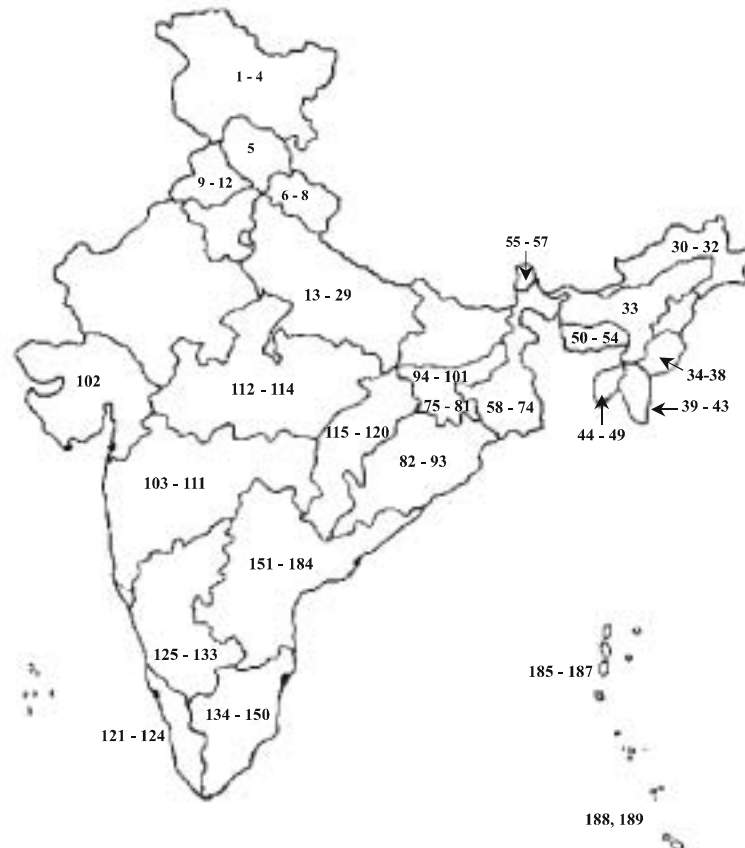
This task has recently been undertaken by molecular geneticists to investigate these complex issues pertaining to Indian populations, by employing markers like autosomal STR's, coding and non-coding regions of the mitochondrial DNA and the non recombining portions of the Y-chromosome. The studies based on these markers have tried to address questions related to the peopling of India by different tribal communities, establishment of caste system in India, gene flow across ethnic groups etc. In this regard, we present a comprehensive review of the studies conducted on Indian populations using different

set of Y-chromosomal markers and also discuss about the issues addressed in relation to the peopling of India by different population groups.

The details of the populations studied by various authors and their geographical affiliation are summarized in Table 1 and Figure 1.

**Table 1: Number of populations studied, for Y-chromosomal polymorphisms, in different geographical regions of India**

<i>Geographic Region</i>	<i>No. of populations studied</i>	<i>References</i>	<i>Denotation in Figure 1</i>
<i>North</i>			
Jammu & Kashmir	4	Kashyap et al. 2006	1 - 4
Himachal Pradesh	1	Kashyap et al. 2006; Sahoo et al. 2006	5
Uttranchal	3	Kashyap et al. 2006	6 - 8
Punjab	4	Kivisild et al. 1999; Basu et al. 2003; Kivisild et al. 2003; Cordaux et al. 2004a	9 - 12
Uttar Pradesh	17	Bhattacharya et al. 1999; Kivisild et al. 1999; Basu et al. 2003; Kivisild et al. 2003; Kashyap et al. 2006; Sahoo et al. 2006; Sengupta et al. 2006; Zerjal et al. 2006	13 - 29
Arunachal Pradesh	3	Cordaux et al. 2004b; Sahoo et al. 2006; Kashyap et al. 2006	30 - 32 33
Assam	1	Cordaux et al. 2004b	
Manipur	5	Basu et al. 2003; Sahoo et al. 2006; Kashyap et al. 2006	34 - 38
Mizoram	5	Kashyap et al. 2006; Sahoo et al. 2006	39 - 43
Tripura	6	Basu et al. 2003; Sengupta et al. 2006	44 - 49
Meghalaya	5	Kumar et al. 2006	50 - 54
Sikkim	3	Kashyap et al. 2006; Sahoo et al. 2006	55 - 57
<i>East</i>			
West Bengal	17	Bhattacharya et al. 1999; Basu et al. 2003; Kivisild et al. 2003; Cordaux et al. 2004a; Kashyap et al. 2006; Sahoo et al. 2006; Sengupta et al. 2006	58 - 74
Bihar	7	Sahoo et al. 2006; Kashyap et al. 2006	75 - 81
Orissa	12	Bhattacharya et al. 1999; Basu et al. 2003; Kashyap et al. 2006; Kumar et al. 2006; Sahoo et al. 2006; Sengupta et al. 2006	82 - 93
Jharkhand	8	Basu et al. 2003; Kashyap et al. 2006; Kumar et al. 2006; Sahoo et al. 2006;	94 - 101
<i>West</i>			
Gujarat	1	Kivisild et al. 2003; Cordaux et al. 2004a; Kashyap et al. 2006; Sahoo et al. 2006	102
Maharashtra	9	Basu et al. 2003; Kivisild et al. 2003; Cordaux et al. 2004a; Kashyap et al. 2006; Sahoo et al. 2006; Sengupta et al. 2006	103 - 111
<i>Central</i>			
Madhya Pradesh	3	Basu et al. 2003; Sengupta et al. 2006	112-114
Chattishgarh	6	Basu et al. 2003; Kashyap et al. 2006	115-120
<i>South</i>			
Kerala	4	Cordaux et al. 2004a; Kashyap et al. 2006	121-124
Karnataka	9	Cordaux et al. 2004a; Kashyap et al. 2006; Sahoo et al. 2006	125-133
Tamil Nadu	18	Basu et al. 2003; Cordaux et al. 2004a; Kashyap et al. 2006; Sahoo et al. 2006; Sengupta et al. 2006	134-150
Andhra Pradesh	34	Bamshad et al. 2001; Ramana et al. 2001; Kivisild et al. 2003; Cordaux et al. 2004a; Kashyap et al. 2006; Sahoo et al. 2006; Sengupta et al. 2006; Thanseem et al. 2006	151-184
Andaman & Nicobar Islands	5	Thangaraj et al. 2002; Kashyap et al. 2006	185-189



**Fig. 1. Map of India showing the location of the populations studied till date for Y-chromosomal polymorphism**

1. Balti, 2. Argon, 3. Drokpa, 4. Buddhist, 5. HP Rajput, 6. Tharu, 7. aunsari, 8. Bhoksha, 9. Lobanas, 10. Punjab Brahmins, 11. Punjab Scheduled castes, 12. Jat Sikhs, 13. UP Brahmins, 14. Kanyakubj Brahmin, 15. Jaunpur Brahmin, 16. Chamar, 17. UP Rajput, 18. Chaturvedi, 19. Khatri, 20. Kurmi, 21. Thakur, 22. UP Jat, 23. UP Kshatriya, 24. UP Vaishya, 25. UP Sudra, 26. UP Panchama, 27. Tharu, 28. Bhoksha, 29. UP Muslims, 30. Adi / Adi Pasi, 31. Nishi, 32. Apatani, 33. Naga, 34. Manipuri Muslims, 35. Meitei, 36. Garo, 37. Naga, 38. Hmar, 39. Mara, 40. Hmar, 41. Lai, 42. Lusei, 43. Kuki, 44. Chakma, 45. Jamatia, 46. Mog, 47. Mizo, 48. Tripuri, 49. Riang, 50. Maram, 51. Bhoi, 52. Pnar, 53. Khyntiam, 54. WarKhasi, 55. Nepali, 56. Bhutia, 57. Lepcha, 58. Santal, 59. Lodha, 60. Karmali, 61. Maheli, 62. Kora, 63. Rajbhansi, 64. Bagdi, 65. WB Brahmins, 66. WB Kshatriya, 67. Gaud, 68. Bauri, 69. Mahishya, 70. Namasudra, 71. WB Kayasth, 72. Tanti, 73. Agharia, 74. Toto, 75. Bihar Brahmin, 76. Bihar Kayasth, 77. Kurmi, 78. Bihar Rajput, 79. Bihar Yadav, 80. Bhumihar, 81. Baniya, 82. Gope, 83. Karan, 84. Khandayat, 85. Oriya Brahmin, 86. Juang, 87. Saora, 88. Paroja, 89. Kharia, 90. Munda, 91. Santal, 92. Agharia, 93. Tanti, 94. Bhumij, 95. Birhor, 96. Ho, 97. Kharia, 98. Munda, 99. Santal, 100. Oraon, 101- Asur, 102. Gujarat Patel, 103. Chitpavan / Konkanastha Brahmin, 104. Desasth Brahmin, 105. Dhangar, 106. Maratha, 107. Katkari, 108. Mahadeo Koli, 109. Pawara, 110. Madia Gond, 111. Naba-Baudh, 112. Halba, 113. Kamar, 114. Muria, 115. Dheria Gond, 116. Satnami, 117. Teli, 118. Agharia, 119. Oraon, 120. Brahmin / Saryupari Brahmin, 121. Kuruchian, 122. Mullakurunan, 123. Paniyan, 124. Nair, 125. Bhovi, 126. Karnataka Christian, 127. Lingayat, 128. Gowda, 129. Iyengar, 130. Kuruva, 131. Koraga, 132. Yerava, 133. Karnataka Muslims, 134. Kota, 135. Kurumba, 136. Irula, 137. Beta Kurumba, 138. Mullu Kurumba, 139. Jenu Kurumba, 140. Toda, 141. Chakkiliyar, 142. Gounder, 143. Kallar, 144. Pallar, 145. Vanniyar, 146. Ambalakarar, 147. Pallan, 148. Iyengar, 149. Iyer, 150. Vellalar, 151. Chenchu, 152. Koya, 153. Naikpod Gond, 154. Andh, 155. Bagata, 156. Lambadi, 157. Valmiki, 158. Poroja, 159. Siddis, 160. Kattunaikan, 161. Koya Dora, 162. Konda Reddy, 163. Yerukula, 164. AP Brahmin, 165. Kamma / Kamma Chaudhary, 166. Kapu / Kappu Naidu, 167. Komati, 168. Raju, 169. Reddy, 170. AP Kshatriya, 171. AP Vaishya, 172. AP Yadav, 173. Relli, 174. Madiga, 175. Mala, 176. Vizag Brahmin, 177. Peruru Brahmin, 178. Golla, 179. Pardhan, 180. Akutota, 181. Panta, 182. Pokanati, 183. Vanne, 184. Sakunpaksollu, 185. Great Andamanese, 186. Onge, 187. Jarwa, 188. Nicobarese, 189. Shompen

### Y-CHROMOSOME VARIATION

The unique properties of uniparental inheritance (paternal), small effective population size, high polymorphism and absence of recombination of Y-chromosome has permitted the employment of this informative locus by researchers worldwide for investigating the genetic structure and the demographic paternal history of human populations. In recent years, certain major issues pertaining to tribes and castes of India have been addressed by hierarchical typing of the slowly evolving Y-chromosome binary polymorphisms (Y-SNP) in congruence with the rapidly evolving Y-chromosome short tandem repeat (Y-STR) loci.

One of the earlier Y-chromosome studies was conducted by Bhattacharyya et al. (1999) on 125 samples from 10 ethnic groups - 8 castes (Brahmins from north, Brahmins from east, Chamar, Bagdi, Mahishya, Agharia, Rajput and Tanti) and 2 tribes (Lodha and Santal) - inhabiting the eastern (West Bengal and Orissa) and northern (Uttar Pradesh) regions of India. The haplotype distribution among the groups showed that different ethnic groups harbor nearly disjoint sets of haplotypes which indicates that there has been virtually no male gene flow among ethnic groups. AMOVA results revealed that there was significant haplotypic variation between castes and tribes, but non-significant variation among ranked caste clusters. Haplotypic variation attributable to differences in geographical regions of habitat was also found to be nonsignificant. In a Y-SNP study on 204 individuals by Ramana et al. (2001) based on three caste groups (Vizag Brahmins, Peruru Brahmins, Kammas), three tribes (Bagata, Poroja, Valmiki) and an additional group (the Siddis) of Andhra Pradesh of South India, evidence of haplotype sharing across caste-tribe boundaries was observed thereby suggesting the possibility of a recent gene flow between the studied castes and tribes in contrast to the negligible gene flow among ethnic groups reported by Bhattacharyya et al. (1999). Another molecular genetic study, involving autosomal loci, mitochondrial DNA and Y-chromosome, by Bamshad et al. (2001) based on 265 samples from eight different caste populations (Upper castes - Niyogi and Vydiki Brahmin, Kshatriya, Vysya; middle castes - Telega and Turpu Kapu, Yadava; lower castes - Relli, Madiga, Mala) from Andhra Pradesh inferred that the upper castes were more similar to Europeans than to Asians, middle castes were equidistant from the

two groups, and lower castes were most similar to Asians. Also, it was observed that the genetic distance between caste populations and Africans was progressively larger moving from lower to middle to upper caste groups. The study concluded that Indian castes are most likely to be of proto-Asian origin with West Eurasian admixture resulting in rank-related and sex-specific differences in the genetic affinities of castes to Asians and Europeans.

In continuation with the above studies, Kivisild et al. (2003), using mtDNA and Y-chromosome markers, conducted a study on the Chenchu and Koya tribes of Andhra Pradesh and compared the obtained results with six caste groups from different parts of India, as well as with western and central Asians. The study supported the concept that India played a crucial role in the late Pleistocene differentiation of the western and eastern Eurasian gene pools. There was no distinction, in the presence or absence of major clades, between tribal and caste groups. Indians showed the presence of diverse lineages of three major Eurasian Y-chromosomal haplo-groups, C, F and K and the exclusive presence of several subclusters of F and K (H, L, R2 and F\*) in the Indian subcontinent (especially H, L and R2) was consistent with the scenario that the southern route migration from Africa carried the ancestral Eurasian lineages to the Indian subcontinent. Westward migration to western Asia and Europe resulted in reduction of four Y-chromosomal founders (C, D, F and K) to two founders (F and K). After the initial settlement, each continental region developed its region-specific branches of these founders, some of which (eg. HV and TJ lineages) have reached back to India. High frequencies of haplogroup R1a together with higher R1a-associated STR diversity was reported (in Punjab and Chenchu tribe) thereby suggesting a southern and western Asian source for this haplogroup. The major east Asian clade O was virtually absent in tribal and caste populations of India with the exception of the subclade defined by M95 whose frequency is highest among the Austro-Asiatic speakers. Overall the results obtained concluded that Indian tribal and caste populations share considerable Pleistocene heritage, with limited recent gene flow between them. This view was however refuted by Cordaux et al. (2004) who proposed independent origins for the Indian caste and tribal paternal lineages based on 931 Y-chromosomes from 15 tribal

groups sampled from southern India and 12 caste groups from all over India. The study reported significant difference in haplogroup frequency distributions in caste and tribal groups with caste groups being more homogenous for Y-chromosome variation and more closely related to each other and to central Asian groups than to tribal groups of India or any other Eurasian groups. H-M52 and F-M89 were the most frequent haplogroups in tribal groups, which were significantly rare in caste groups, whereas haplogroups R-M17, J-M172, R-M124 and L-M20 were prevalent among caste groups and less frequent among tribal groups. As opposed to that proposed by Kivisild et al. (2003), Cordaux et al. (2004) proposed a central Asian origin of R-M17 given its high frequency in central Asia (~20%-40%), its rarity in west Asia and its absence in East Asia. Also, a central Asian rather than a west Asian origin was suggested for Indian J-M172 chromosomes. In sum, haplogroups that are likely to be of indigenous origin are at higher frequency in tribal groups and haplogroups likely of non indigenous origin exhibit higher frequency in caste groups. Hence it was concluded that paternal lineages of the caste populations are primarily descended from Indo-European speakers who migrated from central Asia ~3,500 years ago, while that of the tribal groups are predominantly of indigenous origin. Also, evidences for a bidirectional male gene flow between caste and tribal groups were provided.

A recent study by Sengupta et al. (2006) however argued against any major influence of central Asia on the pre-existing Indian gene pool. The study employed 69 informative Y-chromosome binary markers and 10 microsatellite markers on a large set of geographically, socially, and linguistically representative groups of South Asia and the Y-chromosome binary haplogroup composition was assessed in 728 samples from 36 populations that included 17 tribal populations from six geographic regions and 18 caste populations and a religious group. The study supported the independent origin model put forth by Cordaux et al. (2004) and suggested a partition between (i) Tibeto-Burman speakers and (ii) Austro-Asiatic and Dravidian speakers based on a median-joining network analysis of O2a-M95 microsatellite haplotypes in Indians. Their data also provided overwhelming support for an Indian origin of Dravidian speakers in contrast to the proto-Elamite hypothesis of their origin suggested

by Quintana-Murci et al. (2004) based on mitochondrial DNA polymorphisms. The clear predominance of haplogroup L1-M76 and the greater microsatellite variance within L1 in Dravidian speakers led them to propose an indigenous origin for the Indian Dravidian speakers.

The Y-chromosomal data of 38 Y-SNP's, on 936 Y-chromosomes representing 32 tribal and 45 caste groups, from another recent study by Sahoo et al. (2006) also suggested a largely South Asian origin of Indian caste communities and argued against any major influx, from north and west of India, of people associated with the spread of agriculture. Their findings do support a local origin of haplogroups F\* and H and a west Asian rather than a central Asian origin of haplogroup J2. The claims for the association of haplogroups J2, L, R1a and R2 with the origin of majority of the caste's paternal lineage from outside India, was rejected. In general, the haplogroup frequency distribution (except for haplogroup O and its subclades) across the country was found to be predominantly driven by geographical, rather than cultural determinants. The data also provided a clear evidence for both large-scale immigration of Tibeto-Burman speakers (bearing O3e chromosomes) and language change of former Austro-Asiatic speakers (of O2a lineage) in northeast Indian region. To understand the genetic affinities among the lower castes and tribal groups of India, Thanseem et al. (2006) recently analyzed the Y-chromosome and mitochondrial DNA of 250 males from three tribal populations (Pardhan, Andh and Naikpod) of southern India and compared the obtained results with the available data from the Indian subcontinent. The most frequent haplogroups among the Indian upper castes belonged to R subclades (R\*, R1 and R2) and that among the lower castes and tribal populations to haplogroup H and the distribution pattern of the major Y-lineages was observed to be similar in tribal and lower caste populations, and distinct from the upper castes thereby suggesting a tribal origin for the Indian lower castes, unaffected by geography. Both the tribes and the lower castes harbor indigenous F\*-M89, H-M52, and O-M95 as their major Y-lineages that also supports the genetic closeness of lower castes with the tribal groups. The presence of west / central Asian lineages (J2, R1 and R2) and its higher STR diversity in most of the tribes suggested its presence in India much before the arrival of Indo-European pastoralists

and the impact of their arrival in the sub-continent was interpreted to be social and political rather than genetic.

### LIMITATIONS OF THE ABOVE STUDIES

The earlier studies that dealt with the issue of caste-tribe gene flow, Bhattacharyya et al. (1999) and Ramana et al. (2001), were not extensive enough to draw any strong conclusions about the male gene flow among the Indian ethnic groups, as the former study was restricted to few representative populations of eastern and northern Indian region and the latter was confined to only some populations from Andhra Pradesh of South India. Indeed, the sampling strategy was the major limitation in most of the Y-chromosome studies conducted so far. These studies lacked systematic sampling of populations and involved only few samples from some populations confined to specific geographic regions. In fact, Kivisild et al. (2003), Cordaux et al. (2004) and many others addressed potent issues on a whole Indian scenario based on few representative populations restricted to a definite region. Albeit recent studies (Sengupta et al., 2006; Sahoo et al., 2006) were based on populations from different linguistic groups inhabiting diverse geographical regions of the Indian subcontinent, the sample size of some studied populations was considerably low, for example, in Chakma, Mog (Sengupta et al., 2006), Mara, Bhutia (Sahoo et al., 2006) less than 10 individuals were only studied. So, in order to get a better insight into the peopling of India by the tribal and caste communities of different ethnic status and varied linguistic families, further studies involving greater number of samples from diverse populations inhabiting vast geographical regions are required. Also, increased number of Y-SNP and Y-STR markers should be typed for a detailed resolution and this would permit identification of some region-specific subsets of lineages as well.

### FUTURE PERSPECTIVES

At this juncture what is interesting to know is what the molecular genetic studies have revealed, whether these results support the earlier findings and in what way these studies provide better clarity about the peopling of Indian subcontinent? Were these results able to address the basic concerns of the Indian population structure; the

dichotomy between the castes and tribes; their antiquity and possible origin affinity and diversity within and between these populations? Overall, the Y-chromosome studies attempted on about 200 populations sampled from different geographic regions support the findings of the classical genetic and anthropological studies that there exist wide genetic differences between tribes and upper castes and insignificant differences between the lower castes and tribes and between closely ranked castes. Studies, in general, are in agreement with the independent origin model put forward by Cordaux et al. (2004) which suggests that caste and tribes have independent origins. Current distribution of Y-chromosomal haplogroups in India was primarily associated with geographic rather than linguistic or cultural determinants. Y-chromosome data analysis suggests a Southeast Asian homeland for the Austro-Asiatic speakers (Mundari group) of India (based on distribution of haplogroup O) and an indigenous Indian origin for the Dravidian speakers (based on the haplogroup L1 frequency and associated microsatellite variance). An East/Southeast Asian origin for the Tibeto-Burman speakers of India was suggested based on the (I) presence of haplogroup O3e which is typically found in East/Southeast Asia at frequencies up to 25% to 35% and is absent from India and elsewhere and (II) extensive Y-STR haplotype sharing between northeast India and East/Southeast Asia. With regard to the caste populations, a South Asian origin for the Indian caste communities with minimal Central Asian influence was proposed based on the absence of certain haplogroups in Indian samples (C3, DE, J\*, I, G, N and O) which covers almost half of the Central Asian Y-chromosomes and the presence of some haplogroups in Indian Y-chromosomes (C\*, F\*, H, L and R2) that is poor in Central Asia.

Thus, as evident from the above findings, we now possess a better insight into the peopling of India by population groups of different ethnic status and linguistic families. However, a far better and vivid picture is expected to emerge in future by a stratified sampling strategy so as to represent large set of well-defined ethnic groups from wider geographical regions. Increased molecular resolution by employment of large set of markers and application of appropriate statistical tools for analysis of the results would assist in resolving the persisting enigma regarding Indian populations. In addition, the population based molecular

genetic studies will also enhance our understanding of the health and disease aspect of tribes and castes in India, leading to the effective control of region and population specific diseases.

## REFERENCES

- Bamshad, M., Kivisild, T., Watkins, W.S., Dixon, M.E., Ricker, C.E., et al.: Genetic Evidence on the Origins of Indian Caste Populations. *Genome Res.*, **11**: 994-1004 (2001).
- Basu, A., Mukherjee, N., Roy, S., Sengupta, S., Banerjee, S., et al.: Ethnic India: A Genomic View, With Special Reference to Peopling and Structure. *Genome Res.*, **13**: 2277-2290 (2003).
- Bhasin, M. K. and Walter, H.: *The Genetics of Castes and Tribes of India*. Kamla-Raj Enterprises, Delhi (2001).
- Bhattachayya, N.P., Basu P., Das, M., Pramanik, S., Banerjee, R., et al.: Negligible male gene flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphisms. *Genome Res.*, **9**: 711-719 (1999).
- Buxton, L.H.D.: *The Peoples of Asia*. London (1925).
- Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A.: *The History and Geography of Human Genes*. Princeton University Press, Princeton, NJ (1994).
- Cordaux, R., Aunger, R., Bentley, G., Nasidze, I., Sirajuddin, S.M., and Stoneking, M.: Independent origins of Indian caste and tribal paternal lineages. *Curr. Biol.*, **14**(3): 231-235 (2004a).
- Cordaux, R., Weiss, G., Saha, N., and Stoneking, M.: The Northeast Indian passageway: A barrier or corridor for human migrations? *Mol. Biol. Evol.*, **21**(8): 1525 - 1533 (2004b).
- Guha, B.S.: The racial affinities of the people of India. In: *Census of India, 1931, part III-Ethnographical*. Government of India Press, Simla (1935).
- Kashyap, V.K., Sahoo, S., Sitalaximi, T. and Trivedi, R.: Deletions in the Y-derived amelogenin gene fragment in the Indian population. *BMC Med. Genet.*, **7**: 37 (2006).
- Kivisild, T., Bamshad, M.J., Kaldma, K., Metspalu, M., Metspalu, E., et al.: Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr. Biol.*, **9**:1331-1334 (1999).
- Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., et al.: The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.*, **72**: 313-332 (2003).
- Kumar, V., Langsith, B.T., Biswas, S., Babu, J.P., Rao, T.N., et al.: Global patterns in Human mtDNA Versus Y-chromosome Variation Caused by Spatial Instability of the Local Cultural Processes. *Plos. Genet.*, **2** (4): 53 (2006).
- Majumder, P.P. and Mukherjee, B.N.: Genetic diversity and affinities among Indian populations. An overview, pp. 255-275, In: *Human Population Genetics*. P.P. Majumder (Eds.). Plenum, New York (1993).
- Malhotra, K.C. and Vasulu, T.S.: Structure of human populations in India, pp 207-233, In: *Human Population Genetics*. P.P. Majumder (Eds.). Plenum, New York (1993).
- Pattanayak, D.P.: The Language Heritage of India, pp. 95-99, In: *The Indian Human Heritage*. D. Balasubramanian and N.A. Rao (Eds.). University Press, Hyderabad (1998).
- Poliakov, L.: *The Aryan Myth*. Basic Books, New York (1974).
- Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., et al.: Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am. J. Hum. Genet.*, **74**(5): 827-845 (2004).
- Ramana, G.V., Su, B., Jin, L., Singh, L., Wang, N., et al.: Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur. J. Hum. Genet.*, **9**: 695 - 700 (2001).
- Rapson, E.J.: People and Languages, pp. 33-57, In: *Cambridge History of India, vol 1, Ancient India*. E.J. Rapson (Ed.). S Chand, Delhi (1955).
- Renfrew, C.: Before Babel: Speculations on the origins of linguistic diversity. *Camb. Archaeol. J.*, **1**: 3-23 (1989).
- Rishley, H.H.: *The People of India*. Thacker Spink, Calcutta (1915).
- Sahoo, S., Singh, A., Himabindu, G., Banerjee, J., Sitalaximi, T., et al.: A prehistory of Indian Y-chromosomes: evaluating demic diffusion scenarios. *Proc. Natl. Acad. Sci. USA.*, **103**(4): 843-848 (2006).
- Sarkar, S.S.: Race and race movements in India, pp. 17-32, In: *The Cultural Heritage of India*. S.K. Chatterjee (Ed.). Ramakrishna Mission Institute of Culture, Calcutta (1958).
- Sengupta, S., Zhivotovskiy, L.A., King, R., Mehdi, S.Q., Edmonds, C.A., et al.: Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central Asian pastoralists. *Am. J. Hum. Genet.*, **78**(2): 202-221 (2006).
- Shaffer, J.G.: Harappan culture: A reconsideration, pp. 41-50. In: *Harappan Civilization: A Contemporary Perspective*. G.L. Possehl (Ed.). American Institute of Indian Studies, Oxford and IBH Publishers, New Delhi, India (1982).
- Thanseem, I., Thangaraj, K., Chaubey, G., Singh, V.K., Bhaskar, L.V., et al.: Genetic affinities among the lower castes and tribal groups of India: Inference from Y-chromosome and mitochondrial DNA. *BMC Genet.*, **7**: 42 (2006).
- Thapar, R.: *A History of India* (vol.1). Penguin, Middlesex (1966).
- Zerjal, T., Pandya, A., Thangaraj, K., Ling, E.Y., Kearley, J., et al.: Y-chromosomal insights into the genetic impact of the caste system in India. *Hum. Genet.*, DOI 10.1007/s00439-006-0282-2 (2006).

**KEYWORDS** Y-Chromosome. Indian Populations. Tribes. Castes. Genetic Diversity. Gene Flow. Origin and Antiquity.

**ABSTRACT** An overall view of the results and findings of the recent Y-chromosome studies among 200 populations, representing castes and tribes inhabiting different geographical regions, have been highlighted. While these studies

support the earlier findings based on anthropological and classical genetic markers about the biological implications of castes and tribes; it further gives some clarity about their antiquity and possible origin. However a better understanding of the antiquity of the Indian populations and affinity between castes and tribes is expected to emerge from the future studies with better sampling strategy and representation of more ethnic groups. The molecular genetic studies have the potential to reveal the variation with regard to health and disease aspects among the regional populations.

**Authors' Address: S. Krithika, Suwendu Maji, T.S. Vasulu,** Biological Anthropology Unit, Indian Statistical Institute, Kolkata 700 108, West Bengal, India.

**Corresponding author: 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@isical.ac.in, vasulu@gmail.com