The Human Genetic History of South Asia

Review

Partha P. Majumder*

South Asia — comprising India, Pakistan, countries in the sub-Himalayan region and Myanmar - was one of the first geographical regions to have been peopled by modern humans. This region has served as a major route of dispersal to other geographical regions, including southeast Asia. The Indian society comprises tribal, ranked caste, and other populations that are largely endogamous. As a result of evolutionary antiquity and endogamy, populations of India show high genetic differentiation and extensive structuring. Linguistic differences of populations provide the best explanation of genetic differences observed in this region of the world. Within India, consistent with social history, extant populations inhabiting northern regions show closer affinities with Indo-European speaking populations of central Asia that those inhabiting southern regions. Extant southern Indian populations may have been derived from early colonizers arriving from Africa along the southern exit route. The higher-ranked caste populations, who were the torchbearers of Hindu rituals, show closer affinities with central Asian, Indo-European speaking, populations.

Introduction

Contemporary humans have spread over the entire globe. This dispersal started from Africa. About 100 thousand years ago (kya), a small group of anatomically modern humans migrated out of Africa. We will never know for sure what initiated this migration. It has been speculated that after evolving in Africa, population sizes of humans grew rapidly, eventually outgrowing their natural resources. They initiated exploration of new territories in search of food. There may also have been other causes for the dispersal, such as behavioral changes [1]. From various types of data — in particular, archaeological and genetic - it is now evident that the descendants of those who migrated out of Africa expanded in number to form most contemporary human populations [2-4]. The early out-of-Africa migrants colonized South Asia, dominated by India, and including Pakistan, Nepal, Bhutan, Bangladesh, Sri Lanka and parts of South-East Asia, especially Myanmar.

As humans expanded and migrated out from Africa, they settled down in different geographical locations in small groups. The process of settling may have taken place serially or in a stepwise fashion. That is, as a group of modern humans moved out of Africa and settled down in a new geographical area, they continued to inhabit the new area for several hundred or several thousand years, until population expansion and concomitant pressure on natural resources forced a group of humans from the new area to colonize yet another uninhabited area. The process continued for thousands of years; today humans occupy

Human Genetics Unit, Indian Statistical Institute, Kolkata 700108, India. *E-mail: ppm@isical.ac.in the entire world [3]. Because of the founding of new populations, and evolutionary accumulation of new genetic variation, it is expected that the genetic distance between the source population (Africa) and a descendant population will increase with increasing geographic distance between them. Indeed, such a relationship has been observed [5]. A conceivable scenario of the process of peopling (the geographical spread of modern humans) is that the group of modern humans who moved out of Africa settled in a new area and underwent population expansion in the new area and then a subgroup moved further on to another area. In other words, serial expansion, fission and movement perhaps resulted in the massive geographical spread of modern humans. Support for this mode of peopling has been obtained from genetic data [3]. The process of local population expansion followed by dispersal to newer areas, coupled with evolution of cultural practices and geographical distance acting as a barrier to human contact, also resulted in the formation of intra-marrying groups; groups within which there was considerable gene exchange among individuals, but between which there was little genetic exchange. The development and evolution of language and culture exaggerated the formation of such groups.

Genetic diversity within a geographical area is a reasonable indicator of the age of the populations' residence in that area. However, it is also dependent on the effective sizes of populations, implying thereby that the assessment of antiquity may not be straightforward. It is now well-established that Africa harbors the highest genetic diversity among continental populations. Notably, the genetic diversity in India, the heartland of south Asia, is second only to that of Africa. Thus, it is likely that India became peopled in one of the early waves of migration that originated in Africa. What remains controversial is whether there were multiple waves of migration from Africa, and what the timing of these migration events was. Lahr and Foley [5,6], for instance, postulated that there were two waves: one via a northern exit route through the Nile valley associated with Upper Paleolithic blade-dominated technology (as found in excavation sites in Israel and Lebanon, both dated between 45 and 50 kya) and another, associated with the simpler Middle Paleolithic technology, via a southern exit route from the Horn of Africa across the mouth of the Red Sea along the coastline of India to southeastern Asia and Australia. The dispersal through the southern exit route was postulated to have taken place substantially earlier (about 70-80 kya) than that through the northern route (about 45 kya) [5]. Archaeological evidence to resolve this controversy has been scanty, primarily because the coastlines of that period have become deeply submerged because of the rapid rise of sea levels [7]. However, some recent archaeological finds from India have been interpreted to reveal that the major route of dispersal was through the southern route [8].

Molecular genetic studies, primarily those based on DNA polymorphisms in mitochondrial DNA (mtDNA) and the non-recombining portion of the Y-chromosome (NRY), have favored a single southern dispersal route. mtDNA signatures of Indian and other Asian populations are all derivatives of M and N lineages, which themselves derive from the L3 lineage now found only in Africa. The M and N lineages probably diverged from L3 shortly after their dispersal from Africa [9]. Y-chromosomal data are also indicative of dispersal through the southern route: C and D Y-chromosomal lineages are found only in the Asian continent and Oceania [10,11] and not in western Eurasia and north Africa. Therefore, some human molecular geneticists have argued that the nature and extent of genetic diversity found in non-African populations cannot be reconciled with multiple dispersal events [12,13].

Estimates of the date of migration from Africa into south Asia (India) based on genetic data have also been variable. Most estimates fall between 45 and 60 kya, but a recent genetic study has pushed this date back by about 10,000 years [14]. Some social scientists have also opined that modern humans reached India as early as about 70 kya [15]. This estimate is based on the nature and dating of stone tools and skeletal fossils found at archaeological sites, but is also based on some interpolation of dates of such materials found in Pakistan and Australia. Estimating the date of a migration event from genomic data of contemporary populations is difficult because of uncertainties about past population sizes, rates of genomic changes and selection pressures. Furthermore, mtDNA and Y-chromosomal DNA markers essentially segregate as a single locus - because of lack of recombination among these loci. Together with their nature of transmission — Y-chromosomes only male to male and mtDNA only through females — this implies a strong impact of genetic drift due to a small effective population size. By contrast, the population genetic behavior of autosomal genetic markers can be quite different. These uncertainties and differences have surely contributed to the variability in estimates of dates of migration events or population divergence.

Social Structure of Extant Populations of South Asia

South Asia comprises the present day countries of India, Pakistan, Bangladesh and Sri Lanka. Irrespective of the route taken by modern humans in the out-of-Africa migration and in spite of the uncertainty about its date, there is no doubt that one of the earliest waves reached India at least 50 kya [16]. There is evidence that in the next 20,000 years, modern humans reached Sri Lanka [15]. Contemporary India is a rich tapestry of largely intra-marrying ethnic populations. These populations belong to a diverse set of culture and language groups. Interestingly, the geographical distribution of the language groups within India is largely non-overlapping. The Dravidian speaking groups inhabit southern India, Indo-European speakers inhabit northern India and Tibeto-Burman speakers are confined to northeastern India. By contrast, the numerically small group of Austro-Asiatic speakers, who are exclusively tribal, inhabits fragmented geographical areas of eastern and central India. Culturally, the vast majority of the people of India belong to either tribal or caste societies. The tribal populations are characterized by their traditional modes of subsistence: hunting and gathering, unorganized agriculture, slash and burn agriculture and nomadism (among a limited number of groups). They also have no written form of language and speak a variety of dialects. On the other hand, the Hindu society in India comprises castes, which carry out a wide range of occupations and have written forms of languages. There is overwhelming anthropological and genomic evidence that the tribal populations are the autochthones of India [17,18]. It has been estimated that there are over 400 tribal groups and over 4,000 caste groups in India [19]. Populations belonging to the caste fold have a ranked social order. There are four broad caste groups: The Brahmin or priest caste; the Kshatriya — warrior; Vaisya — trader and husbandman; and, Sudra, the lowest caste [20]. Based on self-perception and others' perception, the caste populations are now ranked as high, middle and low [21]. While the usage of the ranks as 'high', 'middle' and 'low' can appear socially derogatory, "stratification of Indian society reflects and explains a great deal of Indian history, if studied in the field without prejudice" [20]. There is virtually no exchange of genes between tribal populations or between tribal and caste populations. There is also little exchange of genes between castes, primarily due to strict social rules of marriage within the caste system. Social stratification and norms governing mate-exchange between social strata impact on genetic structures of populations. While such stratification and norms exist in many human societies, nowhere are the norms followed with as much rigor as in India. Therefore, human geneticists have studied the genetic structures, similarities and dissimilarities of ranked caste groups in India aiming to shed light on human history, without prejudice. Historical and anthropological studies (e.g., [20,22]) suggest that in the establishment of the caste system in India there have been varying levels of admixture between the tribal people of India and the later immigrants bringing along with them agriculture, pottery and metals from central and west Asia. In other words, castes of different ranks in the contemporary Hindu society putatively have had different degrees of admixture with immigrants into India from central and west Asia. The immigrants from central and west Asia who likely entered India through the north-western corridor, spread to most areas of northern India, but not to southern India. In other words, southern and northern India had differential inputs of genes from central and west Asia. This differential admixture is expected to have differential impacts on the genetic structures of castes of different ranks.

The Enigmatic Original Settlers of South Asia

The Austro-Asiatic speaking populations of India, who are exclusively tribal, show the highest frequencies of the ancient mtDNA lineage M [17]. They also show the highest frequency (about 20%) of sub-lineage M2 [17,23], which has the highest nucleotide diversity within a fast evolving segment (HVS1) of the mtDNA compared to other sub-lineages. Based on these patterns, it has been suggested that the Austro-Asiatic speakers are possibly the earliest settlers of India [17]. Recent results on Y-chromosomal markers provide further support for this inference. The Y lineage O-M95, found in high frequency in India, had originated in the Indian Austro-Asiatic populations around 65 kya [24] note, however, that this age estimate derived from data on length variation of Y-chromosomal short tandem repeat polymorphisms, and may need to be confirmed from data on biallelic markers. These findings are consistent with Renfrew's [25] observation that the present distribution of the Austric language group is owing to the initial dispersal process out of Africa, whereas later agricultural dispersal can account for distribution of the Elamo-Dravidian or Sino-Tibetan languages (to which family Tibeto-Burman languages belong). However, a recent study [26] has found that many Dravidian tribal populations also have M2 frequencies comparable to those of the Austro-Asiatic tribal people,

and the estimated time-depths of the M2 lineage among Austro-Asiatic and Dravidian populations are of similar magnitudes. This and a previous study [27] have emphasized that it would be speculative to tag a population or a set of populations as being descendants of the earliest settlers of south Asia, especially because none of the more ancient lineages can be definitively associated with any specific group of populations, such as populations belonging to a linguistic group.

Geographic, Social and Sex Biases of Gene Flow into India

Analysis of genetic structure has shown that Indian ethnic populations when grouped as tribal versus non-tribal, or by geographical region of habitat, or by linguistic affiliation, have resulted from admixture of four or five ancestral populations [17,18]. In addition to the original African source population, West Asia (by demic diffusion of agriculture) and central Asia have been the major contributors to the Indian gene pool [28], although migrations from these regions have taken place in historical times, perhaps not earlier than 8 kya. The variation of female lineages (mtDNA) in India is rather restricted [29]. By contrast, the variation of male lineages (Y-chromosomal) is very large [30]. This pattern is indicative of sex-biased gene flow into India with more male immigrants than female. Historical data (such as, records of invasions and wars) indicate that predominantly males invaded from central Asia into India [28]. However, the initial view that there had been a large-scale migration from central Asia into India has been significantly modulated. The mtDNA lineage U, which is likely to have arisen in central Asia, has a high frequency in India implying that large-scale migration brought a large number of copies of this lineage into India. However, this lineage was shown to comprise two deep sub-lineages, U2i and U2e, with an estimated split around 50 kya [31]. The sub-lineage U2i is found in high frequencies in India (particularly among tribes) but not in Europe, whereas U2e is found in high frequencies in Europe but not in India (except at very low frequencies among castes, but not tribes [17,31]). Thus, a substantial fraction of the U lineage - specifically, the U2i sublineage - is indigenous to India [30,31].

We note that complete mtDNA genome sequencing has revealed a large number of sequence variants within major haplogroups in Indian populations, many of which, however, occur sporadically [32]. Analysis of these data has indicated a common spread of the root haplotypes of haplogroups M, N and R around 70–60 kya along the southern exit route. The analysis has further revealed that entry of the haplogroup U2 dates after the earliest settlement along the southern route.

Central Asian populations are supposed to have been major contributors to the Indian gene pool, particularly to the northern Indian gene pool, and the migrants had supposedly moved into India through what is now Afghanistan and Pakistan. Using mitochondrial DNA variation data collated from various studies, we have shown [17] that populations of Central Asia and Pakistan show the lowest coefficient of genetic differentiation with the north Indian populations ($F_{ST} = 0.017$), a higher differentiation ($F_{ST} = 0.042$) with the south Indian populations, and the highest ($F_{ST} = 0.047$) with the northeast Indian populations. Northern Indian populations of other geographical regions of India [17,23]. Consistent with the above findings, a recent study [33] using over

500,000 biallelic autosomal markers has found a north to south gradient of genetic proximity of Indian populations to western Eurasians. This feature is likely related to the proportions of ancestry derived from the western Eurasian gene pool, which, as this study has shown, is greater in populations inhabiting northern India than those inhabiting southern India. In general, the Central Asian populations are genetically closer to the higher-ranking caste populations than to the middle- or lower-ranking caste populations [17,23]. Among the higher-ranking caste populations, those of northern India are, however, genetically much closer (F_{ST} = 0.016) than those of southern India (F_{ST} = 0.031). Phylogenetic analysis of Y-chromosomal data collated from various sources yielded a similar picture [17]. Higher-ranking caste populations have been the torch-bearers of the Hindu caste system that was formalized by the Indo-European immigrants. It is likely, therefore, that there was a greater proportion of admixture between higher-ranking caste populations and Indo-Europeans. The fact that high-ranked caste populations inhabiting southern India do not exhibit as much affinity with central Asian populations as those of northern India may be explained by the recent finding [33] that the south Indian, Dravidian speaking, populations may have admixed with north Indian populations bearing ancestral signatures of the western Eurasian gene pool more recently.

High Genetic Differentiation among Indian Populations

Both tribal and non-tribal groups show significant genetic differentiation, whether autosomal [19], mitochondrial [17] or Y-chromosomal [30] markers are used to measure differentiation. However, as mitochondrial and Y-chromosomal markers can essentially be viewed as a single marker because of lack of recombination, autosomal markers are better suited to measure differentiation. The mean value of genetic differentiation (FST) among 55 ethnic groups based on about 400 autosomal markers was 0.03 (0.027 among caste populations and 0.04 among tribal populations). While this value is lower than that (0.14) among continental populations [34–37], it is about an order of magnitude higher than that (0.003) found among European populations [38]. The main reasons for this high level of genetic differentiation found among Indian populations are their evolutionary antiquity and strong founder effects. The FST values are highest for tribal populations inhabiting different geographical regions or belonging to different linguistic groups; genomic differentiation among caste groups is substantially smaller than among tribal groups [17]. There is stronger geographical sub-structuring of upper-caste populations compared with populations of other ranks [17]. The F_{ST} for Y markers are higher than for mtDNA markers, possibly because of social practices (wives moving to their husbands' residence, but not vice versa) that enhance female mobility compared to male mobility [39,40].

Conclusions

South Asia has served as the major early corridor for geographical dispersal of modern humans from out-of-Africa. The likely route of dispersal ran along the coastline of India, from the Horn of Africa and across the mouth of the Red Sea. The search for archaeological and genetic traces of the earliest settlers of this region has not provided any conclusive evidence. The tribal populations of South Asia are older than the non-tribal populations, and these groups show significant genetic differences. Gene-flow into this region has been sex-biased. The contribution of central Asian populations to the Indian gene pool has been low to moderate, but has been higher in north India than in south India. As a result of antiquity, founder effects and strict rules of endogamy, the populations of this region show high genetic differentiation and structure. A recent study [41] indicates that south Asia has also been a major contributor to the gene pool of southeast Asia. With the availability of new genotyping technologies, diversity studies encompassing a large number of populations, both tribal and caste, need to be undertaken at the genome-wide level to validate the inferences of previous studies, and to understand patterns of micro-evolution of populations of this region.

References

- Mellars, P. (2006). Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. Proc. Natl. Acad. Sci. USA 103, 9381– 9386.
- Oestigaard T., Anfinset N., and Saetersdal T., eds. (2004). Combining the Past and the Present: Archaeological Perspectives on Society (Oxford: BAR International Series 1210).
- Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-Sforza, L.L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. USA 102, 15942–15947.
- Rosenberg, N.A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J.K., and Feldman, M.W. (2005). Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet. 1, e70.
- Lahr, M.M., and Foley, R.A. (1994). Multiple Dispersals and Modern Human Origins. Evolutionary Anthropol. 3, 48–60.
- Lahr, M.M., and Foley, R.A. (1998). Towards a theory of modern human origins: geography, demography and diversity in recent human evolution. Yearbook Phys. Anthropol. 41, 137–176.
- Stringer, C. (2000). Palaeoanthropology: Coasting out of Africa. Nature 405, 24–25.
- Mellars, P. (2006). Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. Nature 313, 796–800.
- 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 sapiens from Africa through eastern Africa. Nat. Genet. 23, 437–441.
- Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Told, H.-V., Stepanov, V., et al. (2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am. J. Hum. Genet. 72, 313–332.
- Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Lahr, M.M., Foley, R.A., Oefner, P.J., and Cavalli-Sforza, L.L. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann. Hum. Genet. 65, 43–62.
- Forster, P. (2004). Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. Philos. Trans. R. Soc. London Ser. B 359, 255–264.
- Endicott, P., Metspalu, M., and Kivisild, T. (2007). Genetic evidence on modern human dispersals in South Asia: Y chromosome and mitochondrial DNA perspectives: The World through the eyes of two haploid genomes. In The Evolution and History of Human Populations in South Asia, M. Petraglia and B. Allchin, eds. (Dordrecht, The Netherlands: Springer), pp. 229–244.
- Shi, H., Zhong, H., Peng, Y., Dong, Y.-L., Qi, X.-B., Zhang, F., Liu, L.-F., Tan, S.-J., Ma, R.Z., Xiao, C.-J., et al. (2008). Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. BMC Biol. 6, 45–54.
- Habib, I. (2001). People's History of India Part I: Prehistory (Aligarh, India: Aligarh Historians Society and Tulika Books).
- 16. Cann, R.L. (2001). Genetic clues to dispersal of human populations: Retracing the past from the present. Science 291, 1742–1748.
- Basu, A., Mukherjee, N., Roy, S., Sengupta, S., Banerjee, S., Chakraborty, M., Dey, B., Roy, M., Roy, B., Bhattacharyya, N.P., *et al.* (2003). Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res. 13, 2277–2290.
- Indian Genome Variation Consortium. (2008). Genetic landscape of the people of India: a canvas for disease gene exploration. J. Genet. 87, 3–20.
- Singh, K.S. (1992). People of India: An Introduction (Calcutta: Anthropological Survey of India).
- Kosambi, D.D. (1964). The Culture and Civilisation of Ancient India in Historical Outline (New Delhi: Vikas Publishing House Pvt. Ltd.). (Reprint 1991).
- Singh, K.S. (1998). The people of India: Diversities and linkages. In The Indian Human Heritage, D. Balasubramanian and N. Appaji Rao, eds. (Hyderabad: Universities Press (India) Ltd.), pp. 78–86.

- 22. Bhasin, M.K., and Walter, H. (2001). Genetics of Castes and Tribes of India (Delhi: Kamla-Raj Enterprises).
- Bamshad, M., Kivisild, T., Watkins, W.S., Dixon, M.E., Ricker, C.E., Rao, B.B., Naidu, J.M., Prasad, B.V.R., Reddy, P.G., Rasanayagam, A., et al. (2001). Genetic evidence on the origins of Indian caste populations. Genome Res. 11, 994–1004.
- Kumar, V., Reddy, A.N., Babu, J.P., Rao, T.N., Langstieh, B.T., Reddy, A.G., Singh, L., and Reddy, B.M. (2007). Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. BMC Evol. Biol. 7, 47.
- Renfrew, C. (1992). Archaeology, genetics and linguistic diversity. Man 27, 445–487.
- Satish, K., Padmanabham, P.B.S.V., Rajasekhara, Ravuri, R., Uttaravalli, K., Koneru, P., Mukherjee, P.A., Das, B., Kotal, M., Xaviour, D., *et al.* (2008). The earliest settlers' antiquity and evolutionary history of Indian populations: evidence from M2 mtDNA lineage. BMC Evol. Biol. 8, 230.
- Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T., 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.
- Thapar, R. (2003). Early India: From the origins to AD 1300 (Berkeley, CA: University of California Press).
- Roychoudhury, S., Roy, S., Dey, B., Chakraborty, M., Roy, M., Roy, B., Ramesh, A., Prabhakaran, N., Usha Rani, M.V., Vishwanathan, H., et al. (2000). Fundamental genomic unity of ethnic India is revealed by analysis of mitochondrial DNA. Curr. Sci. 79, 1182–1192.
- Sengupta, S., Zhivotovsky, L.A., King, R., Mehdi, S.Q., Edmonds, C.A., Chow, C.-E.T., Lin, A.A., Mitra, M., Sil, S.K., Ramesh, A., *et al.* (2006). Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central Asian pastoralists. Am. J. Hum. Genet. 78, 202–221.
- Kivisild, T., Bamshad, M.J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., Laos, S., Parik, J., Watkins, W.S., Dixon, M.E., *et al.* (1999). Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr. Biol. 9, 1331–1334.
- Palanichamy, M.G., Sun, C., Agrawal, S., Bandelt, H.-J., Kong, Q.-P., Khan, F., Wang, C.-Y., Chaudhuri, T.K., Palla, V., and Zhang, Y.-P. (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.
- Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian population history. Nature 461, 489–495.
- The International HapMap Consortium. (2003). The International HapMap Project. Nature 426, 789–796.
- Tishkoff, S.A., and Verrelli, B.C. (2003). Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu. Rev. Genomics Hum. Genet. 4, 293–340.
- Watkins, W.S., Rogers, A.R., Ostler, C.T., Wooding, S., Bamshad, M.J., Brassington, A.M., Carroll, M.L., Nguyen, S.V., Walker, J.A., Prasad, B.V., et al. (2003). Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. Genome Res. 13, 1607–1618.
- Watkins, W.S., Prasad, B.V., Naidu, J.M., Rao, B.B., Bhanu, B.A., Ramachandran, B., Das, P.K., Gai, P.B., Reddy, P.C., Reddy, P.G., *et al.* (2005). Diversity and divergence among the tribal populations of India. Ann. Hum. Genet. 69, 680–692.
- Lao, O., Lu, T.T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., Balascakova, M., Bertranpetit, J., Bindoff, L.A., Comas, D., *et al.* (2008). Correlation between genetic and geographic structure in Europe. Curr Biol. *18*, 1241–1248.
- Bamshad, M.J., Watkins, W.S., Dixon, M.E., Jorde, L.B., Rao, B.B., Naidu, J.M., Prasad, B.V.R., Rasanayagam, A., and Hammer, M.F. (1998). Female gene flow stratifies Hindu castes. Nature 395, 851–852.
- Bhattacharyya, N., Basu, P., Das, M., Pramanik, S., Banerjee, R., Roy, B., Roychoudhury, S., and Majumder, P.P. (1999). Negligible gene-flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphisms. Genome Res. 9, 711–719.
- The HUGO Pan-Asian SNP Consortium. (2009). Mapping human genetic diversity in Asia. Science 326, 1541–1545.