Complex genetic origin of Indian populations and its implications

RAKESH TAMANG[1], LALJI SINGH[1,2,3] and KUMARASAMY THANGARAJ[1],*

[1]CSIR-Centre for Cellular and Molecular Biology, Hyderabad 500 007
[2]Genome Foundation, Hyderabad 500 007
[3]Banaras Hindu University, Varanasi 221 005

*Corresponding author (Fax, +91-40-27160311; Email, thangs@ccmb.res.in)

Indian populations are classified into various caste, tribe and religious groups, which altogether makes them very unique compared to rest of the world. The long-term firm socio-religious boundaries and the strict endogamy practices along with the evolutionary forces have further supplemented the existing high-level diversity. As a result, drawing definite conclusions on its overall origin, affinity, health and disease conditions become even more sophisticated than was thought earlier. In spite of these challenges, researchers have undertaken tireless and extensive investigations using various genetic markers to estimate genetic variation and its implication in health and diseases. We have demonstrated that the Indian populations are the descendents of the very first modern humans, who ventured the journey of out-of-Africa about 65,000 years ago. The recent gene flow from east and west Eurasia is also evident. Thus, this review attempts to summarize the unique genetic variation among Indian populations as evident from our extensive study among approximately 20,000 samples across India.

1. Introduction

Modern humans originated in Africa about 200,000 years before present (ybp) (Cavalli-Sforza 1998; Yotova *et al.* 2007). During the period 55,000 to 85,000 ybp, they started moving out of Africa. Based on the geological findings, Scholz *et al.* (2007) reported that during the East African mega droughts between 135,000 and 75,000 ybp, the water volume of the Lake Malawi reduced by at least 95%. The timing of this mega drought coincided with the timing of migration of modern humans out-of-Africa. It is well established that a relatively small group of modern humans ventured out-of-Africa through the southern coastal route to colonize the Middle East, India, Southeast Asia, Australia (figure 1) and subsequently the other parts of the globe. This indicates that India served as one of the important corridors for human migration. Modern human remains in Sri Lanka dating back to the late Pleistocene (55,000 to 25,000 ybp) have been reported by Kennedy *et al.* (1987). Misra (2001) has reported that by the middle Paleolithic period (50,000 to 20,000 ybp) humans

appear to have spread to different parts of the Indian sub-continent. Chamyal *et al.* (2011) showed the evidence of modern human in Orsang river valley, during 50,000 to 30,000 ybp. In order to get better insight on the migratory route taken by modern humans and to understand the peopling of India, several studies have been carried out on various tribes, castes and religious groups in India. Recent studies have shown evidence (stone tools) of early peopling of India by modern humans dating before and after the Toba eruption around 74,000 ybp. The unearthed tools were consistent with contemporaneous *Homo sapiens* tools in Africa (Petraglia *et al.* 2007; Clarkson *et al.* 2009).

2. Human diversity in India

India is considered as a treasure for the geneticists and evolutionary scholars as it is conglomerated with 4,635 anthropologically well-defined populations, among which 532 are tribes, including 72 primitive tribes (36 hunters and gatherers). They differ from each other with respect to their
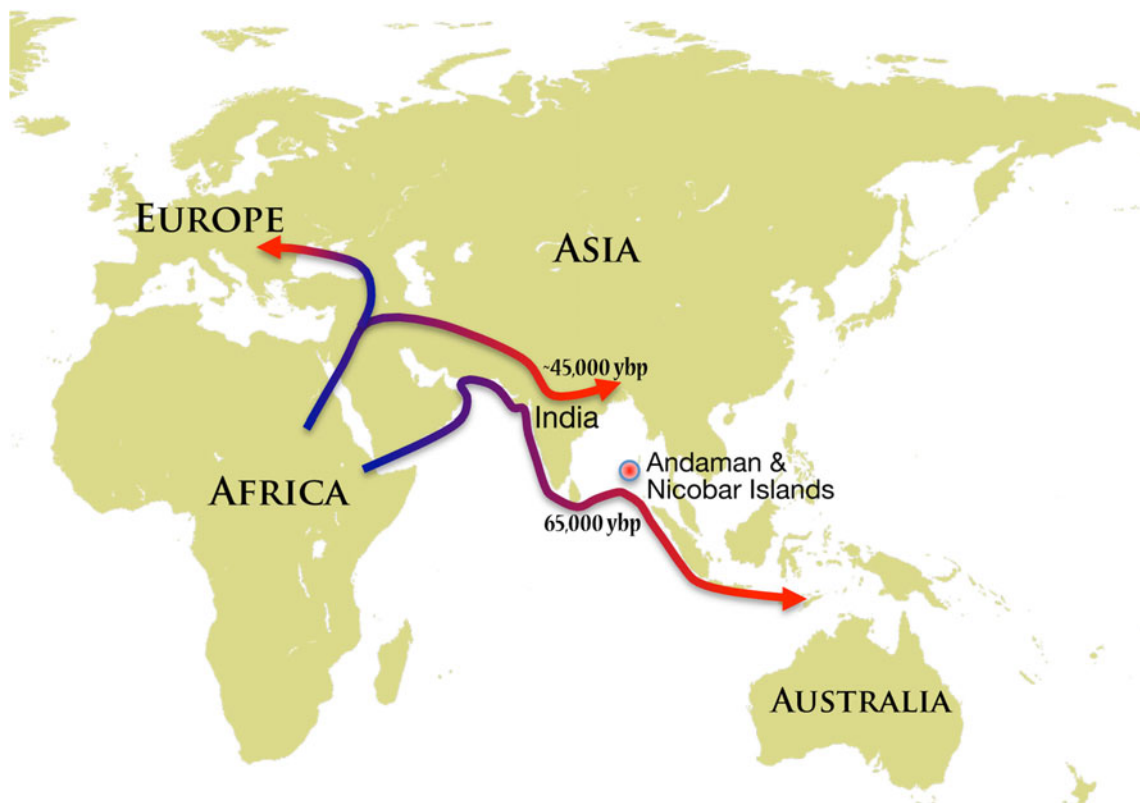
Figure 1.   Possible routes of modern human migrations to Indian subcontinent.

language, social structure, dress and food habits, marriage practices, physical appearance and genetic architecture. India harbours a variety of geographical realms that give refuge to diverse humans and a verity of microbes, plants and animals. In India, four major language families are spoken such as Indo-European, Dravidian, Austroasiatic and Tibeto-Burman. In addition, India has enigmatic Andaman and Nicobar Islanders, whom we predicted as the descendants of early group of modern humans (Thangaraj *et al.* 2003, 2005; Metspalu *et al.* 2004; Macaulay *et al.* 2005) (figure 1).

3.   Social stratification

The caste system was a typical features of the Hindu society and it divided Hindus into four categories viz. Brahmins, Kshatriyas, Vaishyas and Sudras. Brahmins were primarily involved in teachings and performing rituals, Kshatriyas were rulers and defended the territory, Vaishyas were businessmen and the Sudras served as the labourers. Further, each caste is subdivided into subcastes and subcastes into multiple Gotras. The caste system became the governing factor of all socio-religious and economic activities of people. The tribes remained isolated from the other groups and

occupied relatively remote places. Several religious communities build up in mainland India during the course of time due to several waves of migrations from different directions. The rise of the majority of religious groups was basically due to cultural adaptations.

4.   Genetic affinities of Indian populations

4.1   *The era of ancestry makers*

A study using allelic distribution among five major blood-groups among 15 populations (Cavalli-Sforza and Edwards 1967) was the first molecular biological attempt to understand human prehistory. Later, protein and enzyme polymorphisms were used in building human genetic relationships (Nei and Roychoudhury 1982). With the help of the classical markers, Papiha (1996) suggested a distinct demarcation between Indian caste and tribal populations. Later, Y-chromosome and mitochondrial DNA (mtDNA) markers became efficient and handy tools in understanding human phylogeny. mtDNA variations are widely used in deciphering the maternal lineages as it is passed intact from mother to her sons and daughters. Earlier, variations in HVSI

and HVSII regions of mtDNA were extensively used, but soon the sequencing of complete mtDNA came into practice. Y-chromosome markers (STRs and SNPs) are located in the non-recombining region of Y-chromosome (NRY) and can preserve paternal history. Very recently, there has been an increasing use of hundreds of thousands of autosomal SNPs to deduce population structure (Reich *et al.* 2009; Xing *et al.* 2009; Behar *et al.* 2010; Chaubey *et al.* 2011; Metspalu *et al.* 2011; Shah *et al.* 2011).

4.2 *Worldwide scenario*

Y-chromosome and mtDNA markers have been extensively used to infer peopling of different continents/countries and to trace the maternal and paternal lineages of different populations (Hammer *et al.* 1997; Dipierri *et al.* 1998; Kivisild *et al.* 1999, 2004; Qamar *et al.* 1999; Thangaraj *et al.* 1999, 2003, 2005, 2006a; Richards *et al.* 2000; Bamshad *et al.* 2001; Metspalu *et al.* 2004; Rootsi *et al.* 2004; Zhivotovsky *et al.* 2004; Goncalves *et al.* 2005; Bandelt and Kivisild 2006; Hudjashov *et al.* 2007; Underhill and Kivisild 2007; Behar *et al.* 2008; Chaubey *et al.* 2008, 2011). The most accepted model for human origin and migration is known as 'out-of-Africa', which suggests origin of modern human in Africa and subsequent migration and expansion to different continents (Stringer and Andrews 1988); through southern coastal route during 60,000 to 85,000 ybp (Thangaraj *et al.* 2005; Metspalu *et al.* 2004; Macaulay *et al.* 2005) (figure 1). The southern coastal route hypothesis is based on a fact that a small group of modern human on crossing fertile crescent entered India followed by their entry to southeast Asia and subsequently (50,000 to 60,000 ybp) to Australia and rest of the world (Metspalu *et al.* 2004; Thangaraj *et al.* 2005; Macaulay *et al.* 2005). Recently, the early peopling of Europe has been dated approximately 45,000 ybp and many more corrections on the previous dating have been put forward (Callaway 2012).

4.3 *Indian scenario*

The DNA-based studies on Indian populations began during early 1990s. However, some of the initial studies dealt with populations, which are neither anthropologically well-defined nor were really representative Indian populations (Semino *et al.* 1991; Passarino *et al.* 1992; Soodyall and Jenkins 1992; Barnabas *et al.* 1996). Mountain *et al.* (1995) were probably the first, who tried to deal with demographic history of India, based on sequencing of the mitochondrial control (D-loop) region. In a study dealing with the 9 bp deletion located in the mitochondrial genome among a number of tribal and caste populations of southern India, Watkins *et al.* (1999) suggested multiple origin of 9 bp deletion in

southern India, indicating the heterogeneity among the Indians. The traces of socio-cultural, linguistic physio-graphical boundaries and evolutionary forces leading to diversity are well documented in the recent studies. The most accepted and proven view on Indians is that peopling of India is very ancient along with recent gene flow from west and east Eurasia (Kivisild *et al.* 1999; Bamshad *et al.* 2001; Misra 2001; Basu *et al.* 2003; Thangaraj *et al.* 2003, 2005, 2006a, b, 2010; Sengupta *et al.* 2006; Eaaswarkhanth *et al.* 2010; Underhill *et al.* 2010; Chaubey *et al.* 2008, 2011; Chandrasekar *et al.* 2009). We have performed very extensive study among approximately 20,000 Indians including tribes, castes and religious communities using uniparentally inherited Y-chromosome and mitochondrial DNA (mtDNA), and biparentally inherited autosomes.

5. The descendants of early modern humans in India

The Andaman and Nicobar Islands is located in southeast of the Indian subcontinent. It is inhabited by the enigmatic tribes who have remained isolated from the rest of the world for centuries. There are two distinct tribes in the Andaman and Nicobar Islands: (i) those who share physical features with African pygmies and other Asian Negrito people, and (ii) Mongoloid people – those who share physical features with the Southeast Asians. The precise origins of these isolated tribes have been topic of speculation among the scholars of different fields (Abbi 2006; Kumar *et al.* 2006; Chaubey *et al.* 2011). Therefore, we performed an extensive study to understand their origin and evolution using Y-chromosome and mtDNA markers. We sequenced the complete mtDNA genomes of Onges, Great Andamanese and Nicobarese and identified several novel mutations defining two new deep-rooted mtDNA haplogroups, M31 and M32, which are not found anywhere else in the world (Thangaraj *et al.* 2005). We predicted that the Onges and Great Andamanese are the descendants of the early group of modern humans venturing out-of-Africa journey *via* southern coastal route about 65,000 ybp (figure 1). Thus, with this evidence, we proposed an alternate theory of migration; from Africa *via* India to Southeast Asia and Australia, and subsequently to the rest of the world. We took a massive search of these haplogroups (M31 and M32) among the mainland Indians and found none to harbour it. Interestingly, two individuals from Rajbhansi (West Bengal) were found to harbour M31 haplogroup (Palanichamy *et al.* 2006), but the M31 haplogroup was absent among 107 samples of our survey. Additionally, recent studies using high density markers among Indian populations revealed that the Onges are unique (Barik *et al.* 2008; Reich *et al.* 2009).

6. Genetic structure of Indian population

6.1 *Native populations*

India is inhabited by autochthonous as well as migrant populations. The advancement from low-resolution markers to high-throughput whole genome sequencing has enabled us to better understand their complex genetic architecture. In spite of these advancements, studies using high-density markers in India are very limited. Therefore, very few groups have taken rigorous efforts to understand genetic structure of Indian populations using high-density markers, hoping to solve the complexities of Indian genetic structure and its implications. We performed an extensive investigation among various Indian populations using Affymetrix (SNP 6.0) array (Reich *et al.* 2009). We studied a total of 132 individuals from 25 diverse groups representing five major language groups. With powerful statistical analyses, we predicted that the present-day Indian populations probably originated from a relatively small group of isolated ancestors. We predicted the existence of two ancestral groups in the pre-historic India: an 'ancestral North Indian (ANI)', which shared genetic affinity with the populations of the Middle East, Central Asia and Europe (30 to 70%), and an 'ancestral South Indian (ASI)', which has no relation with any population outside India (figure 2) (Reich *et al.* 2009). The present-day Indian populations are the admixture of both ANI and ASI (figure 2). The ASI ancestry may no longer exist in mainland India. The indigenous Andaman Islanders are unique in being only ASI-related groups without ANI ancestry (Reich *et al.* 2009). Further, complete sequencing of the mtDNA has revealed many deep rooted autochthonous haplogroups in India (Thangaraj *et al.* 2003, 2006a, 2009; Chaubey *et al.* 2008; Sharma *et al.* 2012).

6.2 *Waves of immigration*

Beside the indigenous populations, India is also inhabited by several recently migrated populations such as, Siddi, Muslim and the Jews. We have carried out extensive genetic studies among these groups and are well corroborated with their arrival history.

The arrival of Muslim population is well documented. The first Muslims kingdom in India was established in Sindh during 711 AD (Schimmel 1982; Robb 2002). The Turkic Kingdom was established in Delhi during the 13th century, and in the 16th century, the Mughal Empire was established in India. Our study among Indian Muslims inhabited in different parts of India showed their genetic affinity with indigenous non-Muslim populations along with a small frequency of the Middle East ancestry.

Therefore, the spread of Muslims in India was mainly due to cultural adaptation (Eaaswarkhanth *et al.* 2009, 2010).

The Siddis have typical African features and are mainly inhabited in Gujarat, Karnataka and Andhra Pradesh. They were brought to India as slaves and soldiers by Portuguese traders and sold to the Nawabs and the Sultans of India. A high-resolution study among Indian Siddis was lacking. Thus, we screened the Indian Siddis using mitochondrial, Y-chromosomal and autosomal markers and showed a combination of ancestries (i.e. 70% Africans and 30% Indians and Europeans) in the Siddi population (Shah *et al.* 2011). Y-chromosome results revealed that the Siddis are the direct descendants of the Bantu-speakers of sub-Saharan Africa. The signal of unidirectional geneflow from the Indian population to the Siddis confirms the rigidity of the Indian social structure. The effective male population size to arrive India during 17 to 19th century was estimated to be approximately 1,500 individuals (Shah *et al.* 2011). We also estimated that the Siddis were admixed with the local Indian populations since last 8 generations, which is in good agreement with the historical record of their arrival in India (Shah *et al.* 2011).

7. The Indian Austroasiatic speakers

The Austroasiatic language family has 104 million speakers globally. It is widely spoken in central, eastern and north-eastern India. Two major branches of the Austroasiatic language spoken in India are: (i) Munda branch, which is mostly found in eastern, north-eastern and central India, and (ii) Khasi-Aslian branch, mostly spoken in Meghalaya and Nicobar Island (Diffloth 2009). There are two views on the origin and migration of this language (van Driem 2001; Fuller 2007; Kumar *et al.* 2007; Chaubey *et al.* 2011). The first view states Southeast Asia as its place of origin and their subsequent migration to South Asia during the Neolithic (Higham 2003), whereas pre-Neolithic origins and dispersal of this language family from South Asia was hypothesized by the second view (Fuller 2007). A comprehensive study with extensive sampling from both Southeast Asia and South Asia to strongly support the origin of Austroasiatic speakers was lacking. Therefore, we performed an extensive study to trace the precise origin of the Austroasiatic language using uniparentally inherited markers and thousands of SNPs (Chaubey *et al.* 2011). We estimated older coalescent age of 17,000 to 28,000 ybp for the haplogroup M95–O2a, and its higher diversity in Southeast Asia strongly supporting the first of the two views. Thus, the Indian Austroasiatic speakers are derived from Southeast Asia, subsequently followed by sex-specific admixture with local Indians (Chaubey *et al.* 2011).
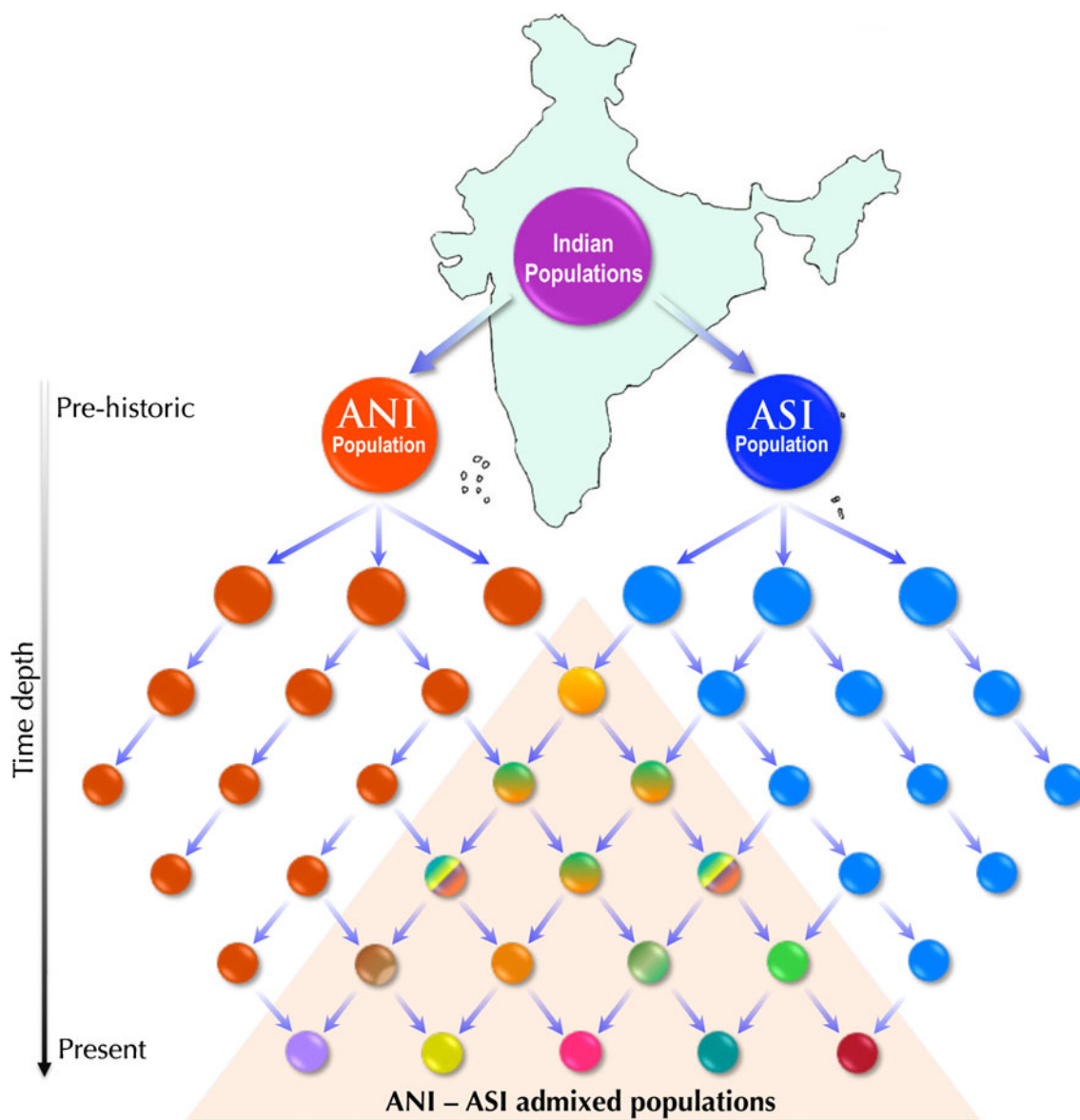
Figure 2. Schematic diagram showing the process of formation of the present-day Indian populations. The hypothetical Ancestral North Indian (ANI) and Ancestral South Indian (ASI) populations got admixed and gave rise to the extant Indian populations, except the Onges (they are unique with only ASI ancestry component).

8. Was there an Aryan invasion?

It is commonly believed that there was an Aryan invasion/ migration to India from the west. However, there is prolonged debate on this topic. It has been well established that various castes and tribal populations of India have a common late Pleistocene maternal as well as paternal ancestry and minor east and west Eurasian ancestries (Kivisild *et al.* 2003; Metspalu *et al.* 2004, 2011; Sahoo *et al.* 2006; Sengupta *et al.* 2006; Chaubey *et al.* 2007, 2008; Reich *et al.* 2009; Shah

et al. 2011; Sharma *et al.* 2012). Most of these studies presumed that the detected west Eurasian genepool may be the Aryan component. Interestingly, both the ANI and ASI ancestry components of the Indian populations are found to harbour higher haplotypic diversity than those predominant in west Eurasia. The shared genetic affinity between the ANI component of northern India and west Eurasia was dated prior to the Aryan invasion (Metspalu *et al.* 2011). These realities suggest the rejection of the Aryan invasion hypothesis but support an ancient demographic history of India.

9. Indian genetic diversity and its medical implications

The genetic heterogeneity among the Indian populations has put forward an immense challenge before the researchers of different fields. The Indian population is governed by various socio-cultural, religious, geographical and linguistic demarcations that ultimately have given birth to strict endogamy practices. Subsequently, this endogamy practices along with evolutionary forces have resulted in higher differences in allele frequency between the groups in India, which has remained intact for thousands of years (Reich *et al.* 2009). Therefore, we have predicted an excess of recessive diseases in India. There are several lines of evidence to support this prediction. We have witnessed various regional and population-specific diseases in many parts of the country, for example, Handigodu disease (Agarwal *et al.* 1994; Badadani *et al.* 2008), Madras motor neuron disease (Nalini *et al.* 2006, 2008), and pseudocholinesterase deficiency among Vyshyas (Rao and Gopalam 1979; Pandit *et al.* 2011), etc. Our recent study using hundreds of thousands of autosomal markers among different ethnic groups in India has showed the evidence of positive selection of *MSTN* and *DOK5* genes (Metspalu *et al.* 2011). These genes have potential implications in lipid metabolism and the etiology of type 2 diabetes only amongst Indians (Metspalu *et al.* 2011). Our other study dealing with *PTPN11* gene, responsible for 50% of the Noonan syndrome (NS), did not find any mutation in *PTPN11* gene in any of the seven NS patients analysed. However, complete mtDNA sequencing revealed all the seven probands and their maternal relatives to be clustered under a major haplogroup R (Rani *et al.* 2010).

In another extensive study, we have shown that the 25bp deletion in the myosin-binding protein-C3 (*MYBPC3*) gene is associated with inheritable cardiomyopathies in India (Dhandapany *et al.* 2009). Forty-five percent of cardiac deaths in India due to sudden heart attack is shown to be caused by this deletion. It is widely distributed (4.5%) across India. Interestingly, it is absent among the Andaman and Nicobar Islanders and Northeastern Indian populations (Dhandapany *et al.* 2009). On extensive screening among worldwide population from 26 countries found this deletion only in the populations of India, Pakistan, Sri Lanka, Indonesia and Malaysia. The time of its origin in these populations was estimated the time of origin of to be about 33,000 years ago (Dhandapany *et al.* 2009).

The unique genetic diversity of the Indian population has frequently yielded surprising results in various association studies. The SNPs/mutations associated with disease among populations of others countries are not usually associated in Indian population. To highlight a few examples: the *POLG*-CAG repeat number variation has been found to be associated with male infertility in several populations but was not associated with Indian infertile men (Rani *et al.* 2009).

Another disagreement with the established association was our study on *SLC11A1* gene polymorphism, which is a strong candidate for infectious disease susceptibility and has been associated with tuberculosis, leprosy and visceral leishmaniasis (VL). Our study on the same gene has shown no association between VL and *SLC11A1* polymorphisms in India (Mehrotra *et al.* 2011).

The admixture event plays a potential role on disease resistance or susceptibility. The admixed Indian Siddis are shown to have remarkable medical and social implications, and has provided a resource for admixture mapping (Narang *et al.* 2011). The Indian Siddis have migrated from Africa and have got admixed with the local Indian populations (Shah *et al.* 2011). A-variant of the *G6PD* gene which is usually found in Africa, is known to give protection against malaria. In spite of the Siddis' African origin, only 10% of the Siddi population was found to harbour the A-variant. This is mainly due to the admixture of Siddis with the local Indian groups ultimately exposing the present-day Siddis to malaria than their African relatives.

10. Conclusions

There are very ancient as well as recent genetic signatures in India. The maternal genepool harbours several deep-rooted lineages, suggesting *in situ* origin of these clades in India. The presence of very ancient lineages in India indicates its major role in early human migration. The genetic studies on Indian population has pointed towards the firm social boundaries, strict endogamy practices and evolutionary forces that have played major role in building the diverse genetic structure of present-day Indian populations. The strict boundaries that prevail at different levels add further to the complexities to the existing diversity, and ultimately raise difficulties in understanding the genetic susceptibility to diseases. A combined inter-disciplinary approach is much needed to explain the precise origin of Indian population as a whole and understanding the disease-associated genetic variants

References

Abbi A 2006 *Endangered languages of the Andaman Islands* (München Lincom Europa)

Agarwal SS, Phadke SR, Phadke RV, Das SK, Singh GK, Sharma JP, Teotia SP and Saxena BN 1994 Handigodu disease a radiological study. A new variety of spondyloepi(meta)physeal dysplasia of the autosomal dominant type. *Skeletal Radiol.* **23** 611–619

Badadani M, Shetty KT, Babu SV and Agarwal SS 2008 Metabolic status of magnesium and ceruloplasmin in Handigodu Joint Disease a variety of spondylo epi (meta) physeal dysplasia. *Clin. Chim. Acta* **395** 170–171

Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, *et al.* 2001 Genetic evidence on the origins of Indian caste populations. *Genome Res.* **11** 994–1004

Bandelt HJ and Kivisild T 2006 Quality assessment of DNA sequence data autopsy of a mis-sequenced mtDNA population sample. *Ann. Hum. Genet.* **70** 314–326

Barik SS, Sahani R, Prasad BV, Endicott P, Metspalu M, Sarkar BN, Bhattacharya S, Annapoorna PC, Sreenath J, Sun D, *et al.* 2008 Detailed mtDNA genotypes permit a reassessment of the settlement and population structure of the Andaman Islands. *Am. J. Phys. Anthropol.* **136** 19–27

Barnabas S, Apte RV and Suresh CG 1996 Ancestry and interrelationships of the Indians and their relationship with other world populations a study based on mitochondrial DNA polymorphisms. *Ann. Hum. Genet.* **60** 409–422

Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, Dey B, Roy M, Roy B, Bhattacharyya NP, *et al.* 2003 Ethnic India a genomic view, with special reference to peopling and structure. *Genome Res* **13** 2277–2290

Behar DM, Metspalu E, Kivisild T, Rosset S, Tzur S, Hadid Y, Yudkovsky G, Rosengarten D, Pereira L, Amorim A, *et al.* 2008 Counting the founders the matrilineal genetic ancestry of the Jewish Diaspora. *PLoS One* **3** e2062

Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, Parik J, Rootsi S, Chaubey G, Kutuev I, Yudkovsky G, *et al.* 2010 The genome-wide structure of the Jewish people. *Nature* **466** 238–242

Callaway E 2012 Archaeology Date with history. *Nature* **485** 27–29

Cavalli-Sforza LL 1998 The DNA revolution in population genetics. *Trends Genet.* **14** 60–65

Cavalli-Sforza LL and Edwards AW 1967 Phylogenetic analysis. Models and estimation procedures. *Am. J. Hum. Genet.* **19** 233–257

Chamyal LS, Maurya DM, Raj R, Juyal N, Bhandari S, Pant RK and Gaillard C 2011 Discovery of a Robust Fossil Homo sapiens in India (Orsang River Valley, Lower Narmada Basin, Gujarat). Possible Continuity with Asian Homo erectus. *Acta Anthropologica Sinica* **2** 158–191

Chandrasekar A, Kumar S, Sreenath J, Sarkar BN, Urade BP, Mallick S, Bandopadhyay SS, Barua P, Barik SS, Basu D, *et al.* 2009 Updating phylogeny of mitochondrial DNA macrohaplogroup m in India dispersal of modern human in South Asian corridor. *PLoS One* **4** e7447

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

Chaubey G, Karmin M, Metspalu E, Metspalu M, Selvi-Rani D, Singh VK, Parik J, Solnik A, Naidu BP, Kumar A, *et al.* 2008 Phylogeography of mtDNA haplogroup R7 in the Indian peninsula. *BMC Evol. Biol.* **8** 227

Chaubey G, Metspalu M, Choi Y, Magi R, Romero IG, Soares P, van Oven M, Behar DM, Rootsi S, Hudjashov G, *et al.* 2011 Population genetic structure in Indian Austroasiatic speakers the role of landscape barriers and sex-specific admixture. *Mol. Biol. Evol.* **28** 1013–1024

Clarkson C, Petraglia M, Korisettar R, Haslam M, Boivin N, Crowther A, Ditchfield P, Fuller D, Miracle P, Harris C, *et al.* 2009 The oldest and longest enduring microlithic sequence in India 35 000 years of modern human occupation and change at the Jwalapuram Locality 9 rockshelter; in *The oldest and longest enduring microlithic sequence in India 35 000 years of modern human occupation and change at the Jwalapuram Locality 9 rockshelter* vol 83 pp 326–348

Dhandapany PS, Sadayappan S, Xue Y, Powell GT, Rani DS, Nallari P, Rai TS, Khullar M, Soares P, Bahl A, *et al.* 2009 A common MYBPC3 (cardiac myosin binding protein C) variant associated with cardiomyopathies in South Asia. *Nat. Genet.* **41** 187–191

Diffloth G 2009 More on Dvaravati Old Mon. *Fourth International Conference on Austroasiatic Linguistics, Mahidol University at Salaya, 29 October 2009*

Dipierri JE, Alfaro E, Martinez-Marignac VL, Bailliet G, Bravi CM, Cejas S and Bianchi NO 1998 Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. *Hum. Biol.* **70** 1001–1010

Eaaswarkhanth M, Dubey B, Meganathan PR, Ravesh Z, Khan FA, Singh L, Thangaraj K and Haque I 2009 Diverse genetic origin of Indian Muslims evidence from autosomal STR loci. *J. Hum. Genet.* **54** 340–348

Eaaswarkhanth M, Haque I, Ravesh Z, Romero IG, Meganathan PR, Dubey B, Khan FA, Chaubey G, Kivisild T, Tyler-Smith C, *et al.* 2010 Traces of sub-Saharan and Middle Eastern lineages in Indian Muslim populations. *Eur. J. Hum. Genet.* **18** 354–363

Fuller D 2007 Non-Human Genetics, Agricultural Origins and Historical Linguistics in South Asia. *Vertebrate Paleobiol. Paleoanthropol.* 393–443

Goncalves R, Freitas A, Branco M, Rosa A, Fernandes AT, Zhivotovsky LA, Underhill PA, Kivisild T and Brehm A 2005 Y-chromosome lineages from Portugal, Madeira and Acores record elements of Sephardim and Berber ancestry. *Ann. Hum. Genet.* **69** 443–454

Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T and Zegura SL 1997 The geographic distribution of human Y chromosome variation. *Genetics* **145** 787–805

Higham CJ 2003 *Languages and farming dispersals austroasiatic languages and rice cultivation* (Cambridge: The McDonald Institute for Archaeological Research)

Hudjashov G, Kivisild T, Underhill PA, Endicott P, Sanchez JJ, Lin AA, Shen P, Oefner P, Renfrew C, Villems R and Forster P 2007 Revealing the prehistoric settlement of Australia by Y

chromosome and mtDNA analysis. *Proc. Natl. Acad. Sci. USA* **104** 8726–8730

Kennedy KA, Deraniyagala SU, Roertgen WJ, Chiment J and Disotell T 1987 Upper pleistocene fossil hominids from Sri Lanka. *Am. J. Phys. Anthropol.* **72** 441–461

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, *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, Metspalu E, Adojaan M, Tolk HV, 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

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E and Villems R 2004 Ethiopian mitochondrial DNA heritage tracking gene flow across and around the gate of tears. *Am. J. Hum. Genet.* **75** 752–770

Kumar V, Langstieh BT, Madhavi KV, Naidu VM, Singh HP, Biswas S, Thangaraj K, Singh L and Reddy BM 2006 Global patterns in human mitochondrial DNA and Y-chromosome variation caused by spatial instability of the local cultural processes. *PLoS Genet.* **2** e53

Kumar V, Reddy AN, Babu JP, Rao TN, Langstieh BT, Thangaraj K, Reddy AG, Singh L and Reddy BM 2007 Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evol. Biol.* **7** 47

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, *et al.* 2005 Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308** 1034–1036

Mehrotra S, Oommen J, Mishra A, Sudharshan M, Tiwary P, Jamieson SE, Fakiola M, Rani DS, Thangaraj K, Rai M, *et al.* 2011 No evidence for association between SLC11A1 and visceral leishmaniasis in India. *BMC Med. Genet.* **12** 71

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, *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

Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, Hudjashov G, Nelis M, Magi R, Metspalu E, Remm M, *et al.* 2011 Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am. J. Hum. Genet.* **89** 731–744

Misra VN 2001 Prehistoric human colonization of India. *J. Biosci.* **26** 491–531

Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M and Cavalli-Sforza LL 1995 Demographic history of India and mtDNA-sequence diversity. *Am. J. Hum. Genet.* **56** 979–992

Nalini A, Yamini BK, Gayatri N, Thennarasu K and Gope R 2006 Familial Madras motor neuron disease (FMMND) study of 15 families from southern India. *J. Neurol. Sci.* **250** 140–146

Nalini A, Thennarasu K, Yamini BK, Shivashankar D and Krishna N 2008 Madras motor neuron disease (MMND) clinical

description and survival pattern of 116 patients from Southern India seen over 36 years (1971–2007). *J. Neurol. Sci.* **269** 65–73

Narang A, Jha P, Rawat V, Mukhopadhyay A, Dash D, Basu A and Mukerji M 2011 Recent admixture in an Indian population of African ancestry. *Am. J. Hum. Genet.* **89** 111–120

Nei M and Roychoudhury A 1982 Genetic relationship and evolution of human races. *Evol. Biol.* **14** 1–59

Palanichamy MG, Agrawal S, Yao YG, Kong QP, Sun C, Khan F, Chaudhuri TK and Zhang YP 2006 Comment on "Reconstructing the origin of Andaman islanders". *Science* **311** 470

Pandit JJ, Gopa S and Arora J 2011 A hypothesis to explain the high prevalence of pseudo-cholinesterase deficiency in specific population groups. *Eur. J. Anaesthesiol.* **28** 550–552

Papiha SS 1996 Genetic variation in India. *Hum. Biol.* **68** 607–628

Passarino G, Semino O, Pepe G, Shrestha SL, Modiano G and Santachiara Benerecetti AS 1992 MtDNA polymorphisms among Tharus of eastern Terai (Nepal). *Gene Geogr.* **6** 139–147

Petraglia M, Korisettar R, Boivin N, Clarkson C, Ditchfield P, Jones S, Koshy J, Lahr MM, Oppenheimer C, Pyle D, *et al.* 2007 Middle Paleolithic assemblages from the Indian subcontinent before and after the Toba super-eruption. *Science* **317** 114–116

Qamar R, Ayub Q, Khaliq S, Mansoor A, Karafet T, Mehdi SQ and Hammer MF 1999 African and Levantine origins of Pakistani YAP+ Y chromosomes. *Hum. Biol.* **71** 745–755

Rani DS, Carlus SJ, Poongothai J, Jyothi A, Pavani K, Gupta NJ, Reddy AG, Rajan MM, Rao K, Chakravarty B, *et al.* 2009 CAG repeat variation in the mtDNA polymerase gamma is not associated with oligoasthenozoospermia. *Int. J. Androl.* **32** 647–655

Rani DS, Dhandapany PS, Nallari P, Govindaraj P, Singh L and Thangaraj K 2010 Mitochondrial DNA haplogroup 'R' is associated with Noonan syndrome of south India. *Mitochondrion* **10** 166–173

Rao PR and Gopalam KB 1979 High incidence of the silent allele at cholinesterase locus I in Vysyas of Andhra Pradesh (S. India). *Hum. Genet.* **52** 139–141

Reich D, Thangaraj K, Patterson N, Price AL and Singh L 2009 Reconstructing Indian population history. *Nature* **461** 489–494

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, *et al.* 2000 Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **67** 1251–1276

Robb P 2002 *A history of India* (Houndmills)

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac L, Pericic M, Balanovsky O, *et al.* 2004 Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in europe. *Am. J. Hum. Genet.* **75** 128–137

Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, Gaikwad S, Trivedi R, Endicott P, Kivisild T, Metspalu M, *et al.* 2006 A prehistory of Indian Y chromosomes evaluating demic diffusion scenarios. *Proc. Natl. Acad. Sci. USA* **103** 843–848

Schimmel A 1982 *Islam in India and Pakistan* (Leiden Brill)

Scholz CA, Johnson TC, Cohen AS, King JW, Peck JA, Overpeck JT, Talbot MR, Brown ET, Kalindekafe L, Amoako PY, *et al.* 2007 East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. *Proc. Natl. Acad. Sci. USA* **104** 16416–16421

Semino O, Torroni A, Scozzari R, Brega A and Santachiara Benerecetti AS 1991 Mitochondrial DNA polymorphisms among Hindus a comparison with the Tharus of Nepal. *Ann. Hum. Genet.* **55** 123–136

Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, Lin AA, Mitra M, Sil SK, 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

Shah AM, Tamang R, Moorjani P, Rani DS, Govindaraj P, Kulkarni G, Bhattacharya T, Mustak MS, Bhaskar LV, Reddy AG, *et al.* 2011 Indian Siddis African descendants with Indian admixture. *Am. J. Hum. Genet.* **89** 154–161

Sharma G, Tamang R, Chaudhary R, Singh VK, Shah AM, Anugula S, Rani DS, Reddy AG, Eaaswarkhanth M, Chaubey G, *et al.* 2012 Genetic affinities of the central Indian tribal populations. *PLoS One* **7** e32546

Soodyall H and Jenkins T 1992 Mitochondrial DNA studies in the South African Indian population. *Gene Geogr.* **6** 127–137

Stringer CB and Andrews P 1988 Genetic and fossil evidence for the origin of modern humans. *Science* **239** 1263–1268

Thangaraj K, Ramana GV and Singh L 1999 Y-chromosome and mitochondrial DNA polymorphisms in Indian populations. *Electrophoresis* **20** 1743–1747

Thangaraj K, Singh L, Reddy AG, Rao VR, Sehgal SC, Underhill PA, Pierson M, Frame IG and Hagelberg E 2003 Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr. Biol.* **13** 86–93

Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA and Singh L 2005 Reconstructing the origin of Andaman Islanders. *Science* **308** 996

Thangaraj K, Chaubey G, Singh VK, Vanniarajan A, Thanseem I, Reddy AG and Singh L 2006a In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. *BMC Genomics* **7** 151

Thangaraj K, Chaubey G, Reddy AG, Singh VK and Singh L 2006b Unique origin of Andaman Islanders insight from autosomal loci. *J. Hum. Genet*. **51** 800–804

Thangaraj K, Nandan A, Sharma V, Sharma VK, Eaaswarkhanth M, Patra PK, Singh S, Rekha S, Dua M, Verma N, *et al.* 2009 Deep rooting in-situ expansion of mtDNA Haplogroup R8 in South Asia. *PLoS One* **4** e6545

Thangaraj K, Naidu BP, Crivellaro F, Tamang R, Upadhyay S, Sharma VK, Reddy AG, Walimbe SR, Chaubey G, Kivisild T and Singh L 2010 The influence of natural barriers in shaping the genetic structure of Maharashtra populations. *PLoS One* **5** e15283

Underhill PA and Kivisild T 2007 Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu. Rev. Genet.* **41** 539–564

Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovsky LA, King RJ, Lin AA, Chow CE, Semino O, Battaglia V, *et al.* 2010 Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur. J. Hum. Genet.* **18** 479–484

van Driem G 2001. Languages of the Himalayas: An ethnolinguistic handbook of the greater Himalayan region with an introduction to the symbiotic theory of language (2 vols.). Leiden: Brill

Watkins WS, Bamshad M, Dixon ME, Bhaskara Rao B, Naidu JM, Reddy PG, Prasad BV, Das PK, Reddy PC, Gai PB, *et al*. 1999 Multiple origins of the mtDNA 9-bp deletion in populations of South India. *Am. J. Phys. Anthropol.* **109** 147–158

Xing J, Watkins WS, Witherspoon DJ, Zhang Y, Guthery SL, Thara R, Mowry BJ, Bulayeva K, Weiss RB and Jorde LB 2009 Fine-scaled human genetic structure revealed by SNP microarrays. *Genome Res.* **19** 815–825

Yotova V, Lefebvre JF, Kohany O, Jurka J, Michalski R, Modiano D, Utermann G, Williams SM and Labuda D 2007 Tracing genetic history of modern humans using X-chromosome lineages. *Hum. Genet.* **22** 431–443

Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, *et al.* 2004 The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* **74** 50–61