# A Genetic Structure of the Early Immigrants (Mukkalathor) of Tamil Nadu as Inferred From Autosomal Loci

**S. Kanthimathi\*, M. Vijaya\*, C. R. Srikumari\*, P. Govinda Reddy\*\*, P. P. Majumder\*\*\* and A. Ramesh\***

*\*Department of Genetics, Dr. ALMPGIBMS, University of Madras, Chennai 600 113, Tamil Nadu, India*
*E-mail: viji@yahoo.co.in*
*\*\*Department of Anthropology, University of Madras, Chennai 600 005, Tamil Nadu, India*
*\*\*\* Anthropology and Human Genetics Unit, Indian Statistical Institute, Kolkata 700 035, West Bengal, India*

**KEYWORDS** Dravidians; *Alu* insertion-deletion polymorphisms; autosomal polymorphic loci; Thevar groups; human migration

**ABSTRACT** Genomic affinity based on eight human – specific polymorphic insertion/ deletion loci was studied in an early immigrant population, Thevar group, of Tamil Nadu, South India. They are traditionally agriculturists, culturally homogenous and endogamous. The seven *Alu* elements *(Alu* APO, *Alu* CD4, *Alu* PV92, *Alu* FXIIIB, *Alu* ACE, *Alu* PLAT, *Alu* D1) and one nuclear insertion of mitochondrial DNA segment (mtNUC) were analyzed for all the DNA samples. All these loci showed high levels of polymorphism in caste populations of Tamil Nadu, thereby reflecting their common ancestry. The significant greater inter-individual variation and the moderate population differentiation probably indicate genetic closeness of these populations. The present study populations were also compared with six other caste populations of Tamil Nadu for which the data were available. Phylogenetic analysis of these populations broadly corresponds to their known ethno-historical affinities.

## INTRODUCTION

India represents one of the most interesting global regions in human evolution, primarily because some of the most ancient out of Africa waves of human migration appear to have passed through India (Cann 2001). Early Eurasians took a route along the Nile across Sinai Peninsula, followed the coast of the Indian Ocean and settled in Northern Australia and Southeast Asia around 60,000 years ago earlier than their settlement in Europe and in Southern Australia (Forster and Matsumura 2005). A second wave of migration (~10,000 years ago) brought in Proto-Dravidian Neolithic farmers from Afghanistan, who were later displaced southwards by a large influx of Indo-European speakers ~3500 years ago into the Indian subcontinent (Majumder 1998). The origin and settlement of the Indian people still remains intriguing, fascinating scientists to explore the impact of these past and modern migrations on the genetic diversity and structure of contemporary populations (Mountain et al. 1995).

Culturally, the Indian populations are stratified as tribes and castes. The contemporary caste populations of India belong to the Hindu religious fold and are arranged hierarchically in Upper, Middle and Lower strata. They predominantly speak languages that belong to the Indo-European or Dravidian families. In addition, there are several communities, who practice different religions other than Hindu religion (Singh 1992).

Population groups inhabiting Tamil Nadu have the distinction of belonging to the Dravidian linguistic family and are predominantly of Australoid ethnicity. Thevar or Mukkalathor is a community of historical antiquity. They constitute 5.5% of the Tamil Nadu State population. Thevar community has three endogamous subgroups (Agamudayar, Maravar and Kallar).

Kallars are known to be the oldest immigrants of Neolithic period with Mediterranean racial elements (Malhotra et al. 1981). Mitochondrial polymorphic marker analysis of the Piramalai Kallars of Madurai, Tamil Nadu - one of the subgroup of the Kallars showed that they might have been the first "Out-of-Africa" migrants, with expansion in the Middle East, subsequent migration to India and would have finally settled in South India (Pitchappan 2002). The Y-chromosomal marker analysis that included the Tamil Nadu populations (Kallar, Yadava and

Sourashtrans), showed the coastal migratory pathway or the Journey of Man out of Africa, to Australia through India (Wells et al. 2002). In view of the significance evidenced from the earlier reports, the Thevar caste group that includes Kallar as one subgroup along with other two subgroups (Maravar, Agamudayar) was included in the present study to screen for autosomal DNA polymorphisms.

Kallar is one of the geographically differentiated endogamous groups of the Thevar community of Tamil Nadu. They are the oldest immigrants of Neolithic period with Mediterranean racial element, who are described as a martial community in the early Chola and Pandya periods (9th century A.D.; Singh 1997). There are six endogamous sub-groups: Esanattu, Gandarava Kottai, Koothappar, Nattar, Periasuriur and Piramalai Kallars. All these groups are placed in Sudra varna. Traditionally, they were described to be semi-agriculturists and semi-warriors. At present, they are mostly agriculturists. They are mainly distributed in Thanjavur and Madurai districts and to a lesser extent in Tiruchi, Ramnad, Tirunelvelli and Pudukottai districts. The size of the population is estimated at over 8 lakhs.

Agamudayar are one of the Mukkalathor. In general, they are agriculturists. Agamudayar of Northern Tamil Nadu use Mudaliar title and those living in Southern parts use Udayar title to their names. They are distributed in the Districts of Thanjavur, Nagapattinam, Tiruchi and Pudukkottai. According to a Maravar historian, there are eleven important subgroups of Maravar: Sembinattu, Kondayam Kottai, Sirutali Katti, Vanniya, Pantara, Karana, Appanurnattu, Agata, Uppukottai, Kuruchi and Sorvaikara Maravar. They are largely distributed in the districts of Tiruchirapalli, Madurai, Kanyakumari, Tirunelveli, Kamarajar, Thanjavur, Pudukottai, Ramanathapuram and Pasumponn Muthuramalinga Thevar Districts. The prescribed exogamous rules are strictly observed i.e., they do not marry within the same clans/gotras (Singh 1992).

The *Alu* family of short interspersed elements is one of the most successful mobile genetic elements, having arisen to a copy number in excess of 500,000 within primate genomes in the last 65 million years. The distribution of these elements varies in different human population groups (Deininger 1981; Batzer et al. 1995). Polymorphic Ya5/8 and Yb8 *Alu* insertions serve as a unique set of nuclear DNA markers for the study of human evolution, as they are stable polymorphisms that are identical by descent. In addition, the ancestral state of each *Alu* insertion is known, facilitating accurate rooting of population networks (Batzer et al. 1995). The frequency distribution of the polymorphic *Alu* elements provides information on the demographic history and migration patterns of human populations (Sherry et al. 1997; Perna et al. 1992).

The present study is an attempt to analyze the extent of diversity of an indigenous Dravidian, Thevar group using eight polymorphic autosomal DNA loci (mtNUC, *Alu* APO, *Alu* CD4, *Alu* PV92, *Alu* FXIIIB, *Alu* ACE, *Alu* PLAT and *Alu* D1). These study populations were also compared with six other caste populations of Tamil Nadu for which the data is available. By and large, it is evident that only progressive accumulation of data can provide a better understanding of evolutionary history of the caste populations and offer a better definition of the genetic landscape of the Indian subcontinent.

## MATERIALS AND METHODS

***Study Populations:*** The three endogamous groups have been sampled from their primary regions of habitat of Tamil Nadu, South India. Information on age, sex, caste and any information concerning their origin, provided by respondents were recorded. Blood samples of healthy and unrelated individuals were collected from Kallar (54) distributed in the areas of Madurai and Thanjavur; Maravar (68) of Kammuthi and Chennai; and Agamudayar (47) of Mannargudi and Thanjavur.

***DNA Isolation and Genotyping:*** After obtaining prior informed consent from the individuals, blood samples (5ml by venipuncture) were collected in sterile EDTA vials. DNA was isolated using the method of Miller et al. (1988). Each DNA sample was screened with respect to eight autosomal insertion/deletion polymorphisms. DNA samples were amplified by polymerase chain reaction (PCR) using locus-specific primers : seven *Alu* elements *(Alu* APO, *Alu* CD4, *Alu* PV92, *Alu* FXIIIB, *Alu* ACE, *Alu* PLAT, *Alu* D1) and one nuclear insertion of mitochondrial DNA segment (mtNUC; Majumder et al. 1999b). All the polymorphic loci studied were genotyped by amplifying 50 – 100 ng of genomic DNA in a standard 30 – cycle three step PCR (Eppendorf Mastercycler Gradient, with heating lid). The PCR

reaction mixture contained the following components in a total volume of 20ml: 10mM Tris HCl (pH 8.4), 50mM KCl, 1.5mM MgCl$_2$, 200 mM dNTPs, 0.5mM each primers, 0.5U *Taq* DNA Polymerase. Amplified PCR products were run on 2 % agarose gel, stained with ethidium bromide, and visualized under UV light.

*Statistical Analysis:* Allele frequencies and their standard errors were computed by the gene counting method. The average heterozygosity was calculated using the estimated allele frequencies for each population. To assess the extent of gene differentiation among the population groups, Nei's (1973) measure of gene diversity was calculated separately for each locus and for all loci considered jointly. The genomic relationships among the populations and dendrograms were constructed by two different methods – UPGMA and Neighbour-joining using DISPAN (Genetic Distance and Phylogenetic Analysis) software (Ota 1993). The relative quantity of gene flow experienced by each population was estimated as described by Harpending and Ward (1982). The Harpending and Ward model assumes random mating within populations. The centroid (*rii*) is defined as the overall mean allelic frequency of the populations. Mathematically the relationship is defined as:

$$rii = (pi - P)2 / P (1- P)$$

where *pi* and *P* are the allele frequencies in population *i* and in the total population, respectively. The theoretical expectation is that there should be linear relationship between heterozygosity and distance from the centroid.

$$i.e., hi = H (1 - rii)$$

where *hi* and H are the heterozygosity of the population *i* and the total population. Populations that have experienced more gene flow than the average fall above the theoretical prediction, whereas populations that have experienced less gene flow than average, fall below the theoretical expectations.

## RESULTS

*Allele Frequencies:* Allelic frequencies for

**Table 1: Allele frequencies at eight polymorphic loci in three caste populations of Tamil Nadu**

| Population | Average Heterozygosity |
|---|---|
| MR | 0.377 ± 0.041 |
| AR | 0.387 ± 0.041 |
| KL | 0.380 ± 0.047 |
| AMB | 0.405 ± 0.053 |
| IYN | 0.449 ± 0.037 |
| IYR | 0.427 ± 0.045 |
| PLN | 0.387 ± 0.059 |
| VAN | 0.400 ± 0.058 |
| VLR | 0.392 ± 0.054 |

the '+' alleles (presence of *Alu* insertions) for the 8 DNA markers are given in Table 1. The *Alu* deletion allele (–) at CD4 locus is human specific and its presence is ancestral state. Chimpanzees, Gorillas, Orangutans and Gibbons are monomorphic for the (+) allele at this locus, thereby indicating that this deletion event occurred likely after the divergence of humans from the great apes about 4 to 6 million years ago (Tishkoff et al. 1996). The frequency of this allele varied about 15% in the study populations. The insertion allele (+) at mtNUC locus is human specific and reported to have occurred before the migration of human populations from Africa. The frequency of this allele varied from 48.5% in Agamudayar to 49.9% in Kallar.

*Average Heterozygosity:* The average heterozygosity was obtained for all loci along with six caste populations of Tamil Nadu for which the data is available (Basu et al. 2003). The values ranged from 37.7% in Maravar to a maximum of 45% in Iyengar (Table 2).

*Genomic Diversity Between Populations:* To determine the amount of genetic differentiation among populations, G$_{ST}$ values (a measure of the interpopulations variability) for each polymorphic locus were determined. The results of gene diversity analysis separately for each locus and for all loci taken together are presented in Table 3. It is seen that except for the *Alu* CD4 locus (0.121) and *Alu* APO (0.299), the total genomic diversity (H$_T$) for each locus was quite high. The values of H$_T$ varied between 0.425 *(Alu* D1) to 0.499 (mtNUC). However, most of the genomic

**Table 2: Gene diversity analysis for individual loci and for all loci considered jointly**

| Populations | mtNUC | Alu ACE | Alu APO | Alu FXIIIB | Alu D1 | Alu CD4 | Alu PLAT | Alu PV92 |
|---|---|---|---|---|---|---|---|---|
| KL | 0.499 | 0.500 | 0.252 | 0.471 | 0.282 | 0.153 | 0.484 | 0.375 |
| AR | 0.485 | 0.494 | 0.258 | 0.486 | 0.258 | 0.156 | 0.477 | 0.440 |
| MR | 0.491 | 0.433 | 0.295 | 0.455 | 0.273 | 0.155 | 0.488 | 0.371 |

**Table 3: Average heterozygosity based on eight loci in nine caste populations of Tamil Nadu**

| Loci | $G_{ST}$ | $H_T$ | $H_S$ |
|---|---|---|---|
| Alu ACE | 0.043 | 0.498 | 0.476 |
| Alu APO | 0.022 | 0.299 | 0.292 |
| Alu CD4 | 0.024 | 0.121 | 0.118 |
| Alu PV92 | 0.034 | 0.476 | 0.460 |
| Alu D1 | 0.060 | 0.425 | 0.399 |
| Alu PLAT | 0.030 | 0.499 | 0.484 |
| Alu FXIIIB | 0.072 | 0.494 | 0.458 |
| mtNUC | 0.031 | 0.499 | 0.484 |
| All Loci | 0.042 | 0.414 | 0.396 |

diversity is attributable to diversity between individuals within the populations (HS). When all loci are jointly considered, 4.2% of the total genomic diversity is attributable to between populations.

***Genomic Affinities among Populations:*** In order to assess the relationship between the populations, genetic distance (DA) was calculated and depicted in a neighbour-joining (NJ) tree. Tree depicts population relationships as a series of bifurcations, which are commonly interpreted as population splits. But, they could also arise from migration instead of shared ancestry. The neighbour-joining tree depicting the population relationships is presented in Figure 1. The nine populations (including 6 other caste populations of Tamil Nadu; Basu et al. 2003) grouped themselves in 5 clusters: (Pallan and Vanniyar), (Iyer and Iyengar), (Ambalakar), (Veerakodi Vellalar) and (Kallar, Maravar, Agamudayar). The Kallar and Agamudayar of the third cluster are more closely related. Amaba-

lakar and Veerakodi vellalar formed separate clusters. These findings were consistent with the known population histories. The relationships of the Tamil Nadu caste populations with the global populations using *Alu* insertion markers are shown in Figure 2. The caste populations, in general, are genetically in between the Caucasoids and Mongoloids. The Thevar group (Kallar, Maravar and Agamudayar) was an exception. They were found to be closer to the people of the United Arab Emirates and Pakistan, and also found intermediate to Africans and Sahul (Australians and New Guineans) populations.

***Gene Flow among Populations:*** With the objective of testing whether in a group of incompletely isolated populations distributed over a geographical space (Wright's island model), observed patterns of genomic diversities were the outcome of the process of drift and migration among the populations or whether these generated by interactions with populations outside the set of populations under consideration, Harpending and Ward (1982) derived a regression of heterozygosity genetic distance from the centroid (rii). This theory assumes a simple linear relationship between the average heterozygosity of a population and genetic distance of the population from the centroid (the overall mean allele frequency of the populations). If a population is receiving genes from outside (these populations used in the analyses) at a higher than average rate, then the heterozygosity will be higher than predicted. A lower than
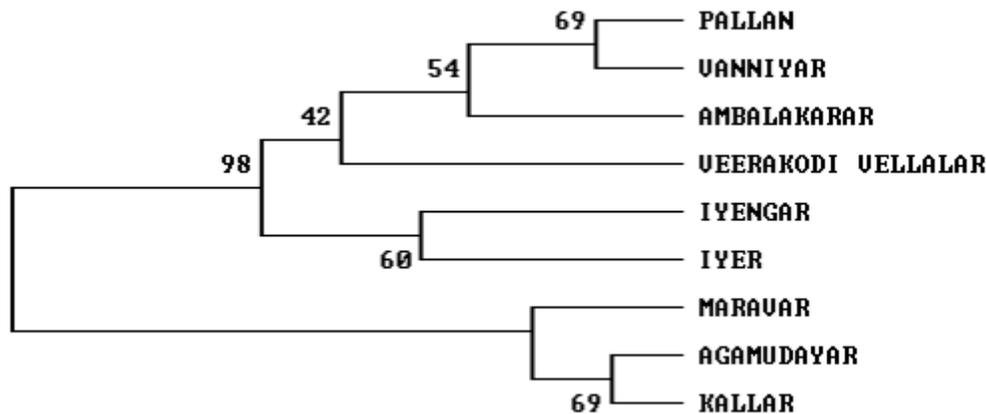


**Fig. 1. Unrooted Neighbour-Joining tree depicting genomic affinity among nine caste populations of Tamil Nadu based on eight human-specific insertion / deletion polymorphisms**
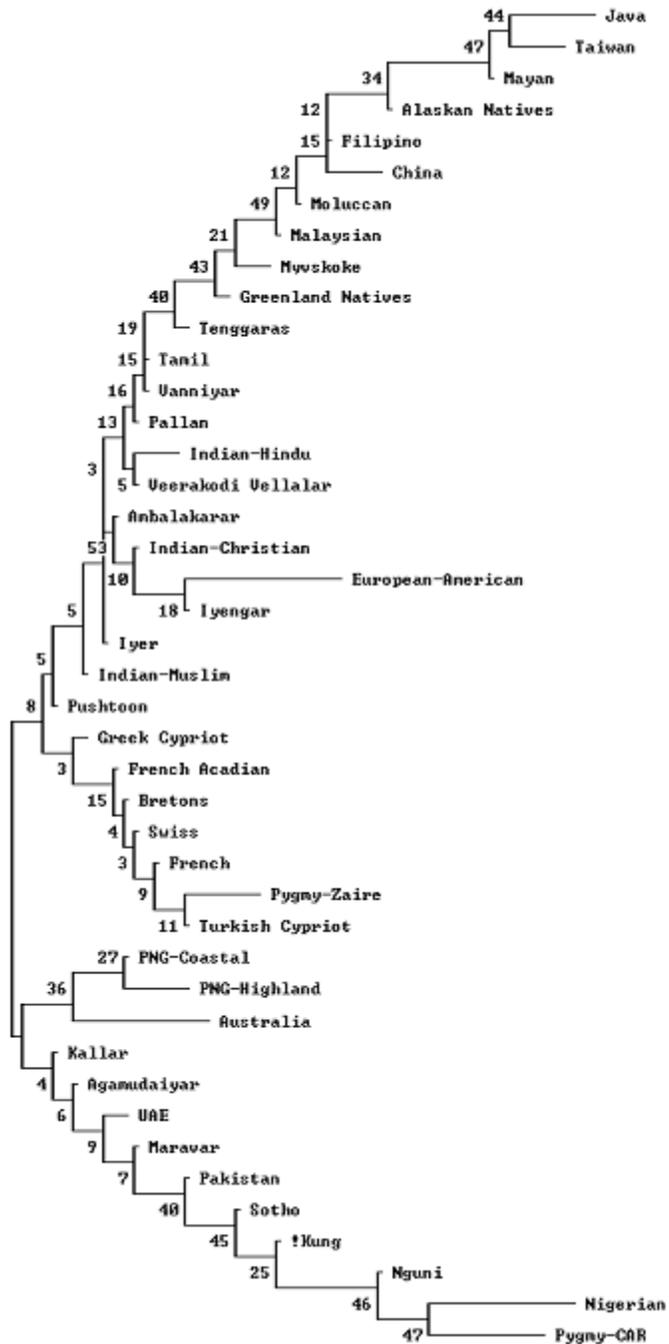
**Fig. 2. Unrooted Neighbour-Joining tree depicting genomic affinity among 43 global populations based on allele frequency at six *Alu* insertion / deletion polymorphisms**
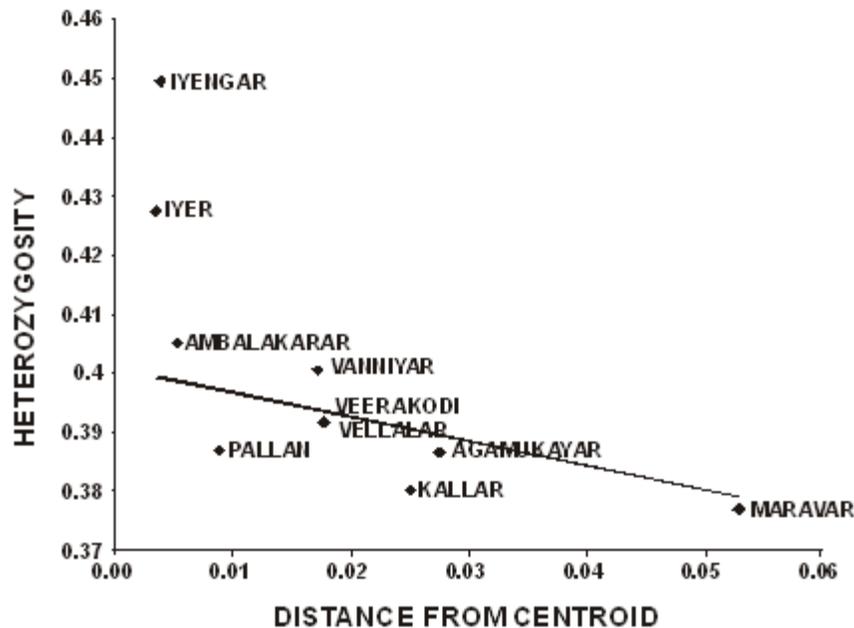
**Fig. 3. Plot of Average heterozygosity vs distance from allele frequency centroid of 9 caste populations of Tamil Nadu based on allele frequency data of eight loci**

predicted heterozygosity implies that the population is more isolated, thereby receiving less gene flow than average.

The R-matrix analysis by pooling the data of the present study with additional six caste groups of Tamil Nadu studied earlier (Basu et al. 2003) has also been carried out. The observed heterozygosity of the nine populations against the distance from the gene frequency centroid is plotted along with the theoretical linear regression line (Fig. 3). The analysis indicated that the Kallar, Maravar, Agamudayar and Pallan have experienced lesser gene flow than predicted, whilst the others are outliers in terms of having higher than predicted heterozygosity. The observed and expected heterozygosity were nearly equal for Veerakodi Vellalar.

## DISCUSSION

Observed variations in the allele frequencies among populations are expected to reflect the sum totality of interplay of evolutionary processes that occurred in the past (Majumder et al. 1999b). As it has been reported in ethnic populations from various parts of the world including India (Stoneking et al. 1997; Kidd et al.

1998; Majumder et al. 1999a, b; Basu et al. 2003), these loci also showed high levels of polymorphism in the caste populations of Tamil Nadu. Since the study populations are generally known to be endogamous, similarities of allele frequency profiles are expected to reflect their common ancestry.

The two Brahmins (Iyer and Iyengar) along with the high rank non-Brahmin formed a separate cluster. The low rank non-Brahmins, Vanniyar and the Harijan – Pallan formed a separate cluster. This is consistent with the earlier studies, that the low rank non-Brahmins are more akin to the Harijans. They are said to have attained a somewhat higher rank than Harijans by accepting the immigrants influence and forsaking beef eating (Balakrishnan and Sanghvi 1981). The high rank non-Brahmin group – Veerakodi Vellalar and low rank non-Brahmin – Ambalakarar are the outliers of the above cluster. The genetic affinities of these populations, thereby, broadly correspond to their known ethno-historical affinities. This present study also confirms earlier findings of Balakrishnan and Sanghvi (1981). Indian populations in general, are genetically in between the Caucasoids and Mongoloids as reported by Majumder et al. (1999) and Viswanathan et al.

(2004). Thevar group, with an exception, was found intermediate to Africans and Sahul (Australians and New Guineans) populations. This may be due to their antiquity.

## ACKNOWLEDGEMENT

## REFERENCES

Balakrishnan V, Sanghvi D 1981. Morphological and genetic distances in Tamil Nadu. In: LD Sanghvi, V Balakrishnan, I Karve (Eds).**:** *Biology of the People of Tamil Nadu.* Bombay: The Indian Society of Human Genetics, Calcutta: The Indian Balakrishnan Anthropological Society, Calcutta, pp. 103-143.

Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, Dey B, Roy M, Roy B, Bhattacharyya NP, Roychoundhury S, Majumder PP 2003. Ethnic India: A genomic view, with special reference to peopling and structure. *Genome Res,* **13**: 2277-2290.

Batzer MA, Rubin CM, Blumberg UH, Hartman MA, Leeflang EP, Stern JD, Bazan HA, Shaikh TH, Deininger PL, Schmid CW 1995. Dispersion and insertion polymorphism in two small subfamilies of recently amplified human *Alu* repeats, *Journal of Molecular Biology,* **247**: 418-427.

Cann RL 2001. Genetic clues to dispersal of human population**s:** Retracing the past from the present. *Science,* **291**: 1742-1748.

Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW 1981. Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J Mol Biol,* **151**: 17-33.

Forster P, Matsumura S 2005. Did early humans go North or South? *Science,* **308**: 965-966.

Harpending, HC, Ward, R 1982. Chemical systematics and human evolution. In: MH Nitechi (Ed.): *Biochemical Aspects of Evolutionary Biology.* Chicago: University of Chicago Press, pp. 213-256.

Kidd KK, Morar B, Castiglione CM, Zhao H, Pakstis AJ, Speed WC, Tamir BB, Lu RB, Goldman D, Lee C, Nam YS, Grandy DK, Jenkins T, Kidd JR 1998. A global survey of haplotype frequencies and linkage disequilibrium at the DRD2 locus, *Human Genetics,* **103:** 211-227.

Majumder PP 1998. People of India; Biological Diversity and Affinities. In: D Balasubramanian, NA Rao (Eds.): *The Indian Human Heritage.* Hyderabad: University Press, pp. 45-59.

Majumder PP, Roy B, Balgir, R, Dash BP 1999a. Polymorphism in the β-globin gene cluster in some ethnic populations of India and their implications on disease. In **:** S Gupta, OP Sood (Eds): *Molecular Intervention in Disease.* New Delhi: Ranbaxy Science Foundation, pp. 75-83.

Majumder PP, Roy B, Banerjee S, Chakaraborty M, Dey B, Mukherjee N, Roy M, Thakurta PG, Sil SK 1999b. Human specific insertion/deletion polymorphism in Indian populations and their possible evolutionary implications, *Euro J Hum Genet,* **7**: 435-446.

Malhotra KC, Balakrishnan V, Karve I 1981. Anthropometric variation in Tamil Nadu. In: LD Sanghvi, V Balakrishnan, I Karve (Eds).: *Biology of the People of Tamil Nadu.* Bombay: The Indian Society of Human Genetics, Calcutta: The Indian Balakrishnan Anthropological Society, Calcutta, pp. 50-74.

Miller SA, Dykes DD, Polesky HF 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research,* **16(3):** 1215.

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

Nei M 1973. Analysis of gene diversity in subdivided populations, *Proc Natl Acad Sci,* USA, **70**: 3321-3323.

Ota T 1993. *DISPAN: Genetic Distance and Phylogenetic Analysis.* Pennsylvania: Pennsylvania State University.

Perna NT, Batzer, MA, Deininger PL, Stoneking M 1992. *Alu* insertion polymorphism: A new type of marker for human population studies. *Human Biology,* **64(5):** 641-648.

Pitchappan RM 2002. Castes, Migration, Immuno-genetics and infectious diseases in South India. *Community Genet,* **5**: 157-161.

Sherry ST, Harpending HC, Batzer MA, Stoneking M 1997. *Alu* evolution in human populations: using the coalescent to estimate effective population size, *Genetics,* **147**: 1977-1982.

Singh KS 1992. *People of India: An Introduction.* Anthropological Survey of India, Calcutta: Seagull Publication.

Singh, KS 1998. *People of India, India's Communities*, National Series Vol. V, Anthropological survey of India, New Delhi: Oxford University Press.

Stoneking, M, Soodyall, H 1996. Human evolution and the mitochondrial genome. *Curr Opin Genet Dev,* **6**: 731-736

Vishwanathan H, Edwin D, Cordaux R, Stoneking M, UshaRani MV, Majumder PP 2004. Genetic structure and affinities among tribal populations of southern India: a study of 24 autosomal DNA markers, *Annals of Human Genetics,* **68:** 128-138.

Wells S 2002. *The Journey of Man, a Genetic Odyssey.* UK: The Penguin Press.