RESEARCH NOTE

Genetic study of scheduled caste populations of Tamil Nadu

M. VIJAYA*, S. KANTHIMATHI and A. RAMESH

Department of Genetics, Dr ALM PGIBMS, University of Madras, Taramani, Chennai 600 113, India

Introduction

The present study aims to describe the genetic structure of the scheduled caste populations in Tamil Nadu state, and to assess their relationships with contemporary people of different socio-economic groups of the state. We have studied eight human-specific indels (insertion/deletion polymorphisms) in DNA samples from five Tamil Nadu caste groups using phylogenetic and principal component analysis (PCA), and compared the data with earlier data on the three other caste groups of the state. The results indicate that scheduled castes are genetically heterogenous, but show closeness with other socio-economic groups of the state.

The social structure of India is dominated by the Hindu caste system. Caste refers to a rigid system of ranked social inequality with significant barriers to mobility or to intimate associations between different strata. 'Caste' comes from the Portuguese word *casta*, meaning lineage, breed, race or category, and is described by one or both of the two ancient Indian concepts of varna and jati (Pillay 1979): 'Studies in the history of India with special reference to Tamil Nadu'.

The caste (varna) system has four ranked categories: priests (Brahmins), warriors (Kshatriyas), farmers and merchants (Vaisyas), and labourers and servants (Sudras). However, the status of the Sudra was actually higher than that of a fifth varna, the Panchama. This fifth varna was added at a later stage to include the so-called untouchables, who were excluded from the other four varna (Elder 1996). Based on socio-economic hierarchy, the varna system can be classified into three castes: upper (Brahmins), middle (Kshatriyas, Vaisyas and Sudras) and lower (Panchama). In contemporary India, the government refers to Panchama or untouchables as scheduled castes.

Scheduled caste members comprised about 139 million people, or more than 16 per cent of the total population of India. About 10.7 million of them are distributed in Tamil Nadu state, the majority being Hindus (Census of India 2001). Major scheduled castes of Tamil Nadu are Pallan, Paraiyan and Chakkiliyar, belonging to the Dravidian linguistic group.

The present study was undertaken to delineate the genetic affinity of scheduled castes among themselves and with other varnas of the Hindu caste system. We have genotyped the study groups using seven human specific *Alu* insertions (Stoneking *et al*. 1997), a mitochondrial DNA insert in the human nuclear genome (Zischler *et al*. 1995), and a deletion of 256 bp of a 285-bp *Alu* element at the *CD4* locus (Edwards and Gibbs 1992). *Alu* elements are stable genetic markers and their frequency distribution provides information on the demographic history and migration patterns of human populations. The knowledge of both the ancestral state and the direction of mutational change facilitate phylogenetic analysis (Batzer *et al*. 1995). Earlier study data of other Tamil Nadu caste groups (upper caste: Iyer and Iyengar, lower caste: Pallan) were used for comparative analysis.

Caste groups

Pallan

The Pallan are an ancient community, engaged extensively in wet land farming and distributed mainly in Thanjavur, Madurai and Ramanathapuram districts (Ramaiah 2004). The Pallan rank themselves highest among the scheduled castes, and number around 2.2 million (Census of India 2001).

Paraiyan

The term Paraiyan is derived from the Tamil word parai meaning drum, as some of them act as drummers at funerals and village festivals. They are also engaged in cultivation, grass cutting and weaving. Fifteenth century literature indicates that the Paraiyans were also engaged in leather processing from dead animals, which is traditionally considered to be defiling and polluting. They are also known as Sambavan,

Sambavar and Samban (Singh 1998), and their number was estimated to be around 1.8 million (Census of India 2001).

Chakkiliyar

Chakkiliyars are also referred to as Arundhatiyar, Madari, Madiga and Pagadai, and represent the lowest strata even among the schedule castes. No literature is available to indicate their historical origin, although it is thought that they might be immigrants from the neighbouring state of Andhra Pradesh, as their mother tongue is Telugu. They are mainly a landless community. Their traditional occupations are sweeping, scavenging and removing dead animals, making and repairing footwear, as well as working as farm labourers (Singh 1998). The size of the population was estimated to be over 0.7 million (Census of India 2001).

Kallar

Kallars are known to be the oldest immigrants of Neolithic period with Mediterranean racial elements. Traditionally, they were described as semiagriculturists and semiwarriors. They are mostly agriculturists, and are mainly distributed in Thanjavur and Madurai districts and to a lesser extent in Tiruchirapalli, Ramanathapuram, Tirunelveli and Pudukkottai districts (Singh 1998). The size of the population was estimated to be over 0.8 million (Second Backward Classes Commission of Tamil Nadu 1989).

Maravar

They are largely distributed in the districts of Tiruchirapalli, Madurai, Kanniyakumari, Tirunelveli, Virudhunagar, Thanjavur, Pudukkottai, Ramanathapuram and Sivaganga districts. Traditionally, they are agriculturists (Singh 1998). The population size was estimated around 0.53 million (Second Backward Classes Commission of Tamil Nadu 1989).

Agamudaiyar

Agamudaiyar of northern Tamil Nadu use the title 'Mudaliar' title whereas those living in southern parts use the title 'Udayar'. They are distributed in the districts of Thanjavur, Nagapattinam, Tiruchirapalli and Pudukkottai (Singh 1998), and number around 0.82 million (Second Backward Classes Commission of Tamil Nadu 1989).

Subjects and methods

The populations selected for the present study includes scheduled castes (lower caste: Paraiyan and Chakkiliyar), and middle castes (Kallar, Maravar and Agamudaiyar). In addition to these, upper caste (Iyer and Iyengar) and lower caste (Pallan) groups, for which comparable data were available and were also included for comparative analysis (Basu *et al*. 2003). All these populations belong to Dravidian linguistic group.

Blood samples were drawn with informed consent, from 266 unrelated individuals belonging to two scheduled caste:

Paraiyan ($n = 50$) and Chakkiliyar ($n = 47$), and three middle caste: Kallar ($n = 54$), Maravar ($n = 68$) and Agamudaiyar ($n = 47$). Information about age, sex and caste provided by the respondents were recorded. Institutional human ethical committee clearance was obtained prior to the collection of blood samples. Paraiyan samples were collected from Cuddalore, Dharmapuri and Chennai, Chakkiliyar samples were from Salem and Erode districts, Kallar samples from Madurai and Thanjavur, Maravar samples from Kammuthi and Agamudaiyar samples from Thanjavur district.

DNA from the peripheral blood leucocytes was isolated using the protocol of Miller *et al*. (1988). DNA samples were screened for eight indel consisting of seven *Alu* insertion sites (*Alu APO*, *Alu CD4*, *Alu PV92*, *Alu FXIIIB*, *Alu ACE*, *Alu PLAT* and *Alu D1*) and one mitochondrial DNA insertion site (*mtNUC*) (Viswanathan *et al*. 2004). These DNA markers were selected by the Department of Biotechnology (DBT), Government of India as part of its Human Genome Diversity (HGD) initiative.

The allele frequencies and their standard errors were computed by gene counting method. Chi-square tests for the departure from Hardy–Weinberg expectations were performed. Average heterozygosity was calculated using the estimated allele frequencies for each population. DISPAN (genetic distance and phylogenetic analysis) software (Ota 1993) was used to compute average heterozygosity and gene diversity (Nei 1973). The software also constructs phylogenetic trees (dendrograms) by using the neighbour-joining method. Population relationships were also analysed via PCA, using SPSS 11.0 version.

Results

Allele frequencies

The *Alu* deletion allele (−) at *CD4* locus is human specific and its presence is considered to be the ancestral state. Chimpanzees, gorillas, orangutans and gibbons are monomorphic for the (+) allele at this locus, thereby indicating that this deletion event probably occurred after the divergence of humans from the great apes about 4 to 6 million years ago (Tishkoff *et al*. 1996). The frequencies of these allele were less than 11% 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 (Zischler *et al*. 1995). The frequency of this allele varied from 30.2% in Chakkiliyar to 61.8% in Iyengar. No significant departure from the Hardy–Weinberg equilibrium was observed ($P < 0.05$).

Average heterozygosity

The average heterozygosity was obtained for all loci along with three caste populations of Tamil Nadu for which the data are available (Basu *et al*. 2003). The values ranged from 37.4% in Maravar to a maximum of 45% in Iyengar. The average heterozygosity values of the study populations

are higher than the other global populations studied, with the exception of African populations (Stoneking *et al.* 1997; Novick *et al.* 1998). Thus, the DNA markers attest that the caste populations of Tamil Nadu exhibit high levels of genomic diversity.

Table 1. Gene diversity analysis for individual loci and for all loci considered jointly.

Loci	H_t	H_s	G_{st}
mtNUC	0.4944	0.4727	0.0400
Alu ACE	0.4994	0.4795	0.0398
Alu APO	0.3370	0.3254	0.0344
Alu FXIIB	0.4997	0.4714	0.0567
Alu D1	0.3994	0.3712	0.0707
Alu CD4	0.1366	0.1341	0.0183
Alu PLAT	0.4959	0.4744	0.0432
Alu PV92	0.4464	0.4265	0.0445
All loci	0.4136	0.3944	0.0464

H_t–Total genomic diversity among the populations.
H_s–Diversity between individuals within population.
G_{st}–Genomic diversity between population.

Genomic diversity between populations

The amount of genetic differentiation among populations, G_{ST} values (a measure of the interpopulation variability) for each polymorphic locus were determined. The results of the gene diversity analysis separately for each locus and for all loci taken together are presented in table 1. It is seen that except for the *Alu CD4* locus (0.018) and *Alu APO* (0.034), the total genomic diversity (H_T) for each locus was quite high. The values of H_T varied between 0.137 (*Alu CD4*) to 0.499 (*Alu FXIIIB*). However, most of the genomic diversity is attributable to diversity between individuals within the populations (H_S). When all loci are jointly considered, 4.6% of the total genomic diversity is attributable between-populations variation.

Genomic affinities among populations

The neighbour-joining tree depicting the population relationships is presented in figure 1. The eight populations (including three other caste populations of Tamil Nadu; Basu *et al.* 2003) grouped in six clusters: (Kallar and Maravar), Agamudaiyar, Chakkiliyar, Paraiyan, Pallan and (Iyer and Iyengar). Upper caste and middle caste populations formed separate clusters, while the scheduled caste populations fall on independent branches.

PCA of the allele frequencies at the eight polymorphic Indel loci was also performed (figure 2). The two principal components, explains about 94% of the allele variance, which provide the most information for a two-dimensional depiction of population relationships.

Discussion

The present study aimed at revealing the genetic relationship of the scheduled caste (lower caste) among themselves and with other socio-economic groups (upper and middle) in Tamil Nadu.

The Phylogenetic analysis illustrated that the clustering pattern is similar to that of the different socio-economic groups, except for the lower caste (figure 1). Upper and middle caste populations showed close affinity among themselves, whereas, the scheduled castes did not show any closeness. These findings demonstrated that the scheduled caste populations of Tamil Nadu are genetically heterogenous.

Tamil Nadu state was colonized by several waves of migrants during different time periods. Based on morphological traits Malhotra *et al.* (1981) recognized the existence of three morphological types among the populations of Tamil Nadu. The Proto–Australoid features were identified in Paraiyan, Chakkiliyar and Mutracha. People with these features, according to them, might have entered the Tamil region between 15,000–20,000 years ago. Paleo–Mediterranean and Mediterranean features were identified in Kallar and Vanniyar. The antiquity of Mediterranean populations has
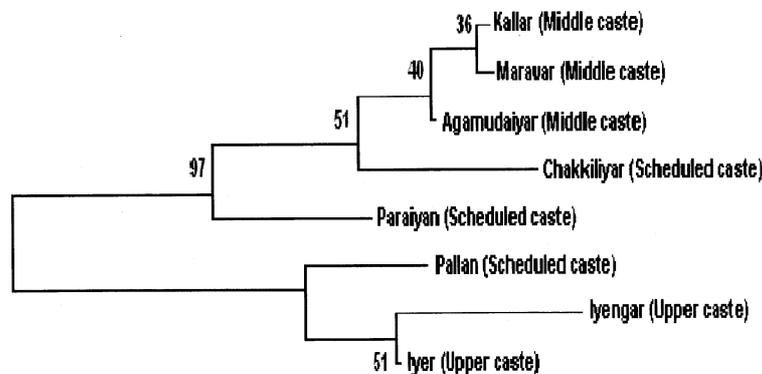

Figure 1. Unrooted neighbour-joining tree depicting genomic relationships among 8 caste populations of Tamil Nadu. Number indicate (%) the fraction of 5000 bootstrap replicates that supported a particular grouping.

been placed 10,000 years ago. Western branchycephalic features identified in Mudaliars, Vellalars, Chettiars, Brahmins and Naidus, might have entered at least 3,000 to 5,000 years ago.

The PCA (figure 2) placed Paraiyan, Chakkiliyar and Kallar in the same quadrant, even though Kallar having distinct morphological features. Although, Pallan of scheduled caste group placed in separate quadrant (figure 2), showed affinity with Brahmins (Iyer and Iyengar) in phylogenetic analysis (figure 1). The genetic affinities of these populations, thus, broadly correspond to their known ethnohistorical affinities. This present study also confirms earlier findings of Balakrishnan and Sanghvi (1981).
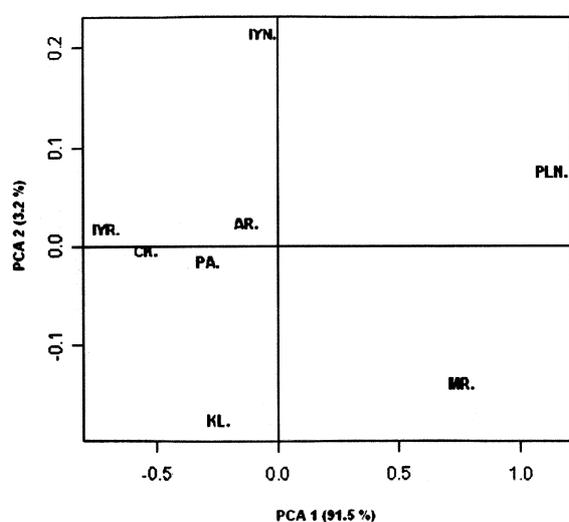

Figure 2. Principal component analysis depicting genomic relationships among eight caste populations of Tamil Nadu. Upper caste (IYN, Iyengar; IXR, Iyer), middle caste(KL, kallai; MR, Maravar; AR, Agamuclaiyar) and lower caste (PA, Paraiyan; PLN, Pallan; CH, chakkiliyar).

Overall, this study showed that the scheduled caste populations of Tamil Nadu are genetically heterogenous, but also showed affinity with contemporary populations of different socio-economic groups.

References

Balakrishnan V. and Sanghvi L. D. 1981 Morphological and genetic distances in Tamil Nadu. In *Biology of the people of Tamil Nadu.* (ed. L. D. Sanghvi, V. Balakrishnan and I. Karve), pp. 103–143 The Indian society of Human Genetics Mumbai, and The Indian Balakrishnan Anthropological society, Kolkata.

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

Batzer M. A., Rubin C. M., Blumberg U. H., Hartman M. A., Leeflang E. P., Stern J. D. *et al.* 1995 Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. *Mol. Biol.* **247**, 418–427.

Census of India 2001 Provisional population totals: rural-urban distribution. Paper 1 of 2001, series 34, Director of census operation, Tamil Nadu.

Edwards M. C. and Gibbs R. A. 1992 A human dimorphism resulting from loss of an *Alu. Genomics* **14**, 590–597.

Elder J. 1996 Enduring stereotypes about South Asia: India's caste system. *Educ. Asia* **1**, 20–22.

Malhotra K. C., Balakrishnan V. and Karve I. 1981 Anthropometric variation in Tamil Nadu In *Biology of Tamil Nadu* (ed. L. D. Sanghvi, V. Balakrishnan and I. Karve), pp. 50–74 Indian society of Human Genetics, Pune, and The Indian Anthropological society, Kolkata.

Miller S. A., Dykes D. D. and Polesky H. F. 1988 A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215.

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

Novick G. E., Novick C. C., Yunis J. Yunis E., Mayolo A. D., Scheer W. D. *et al.* 1998 Polymorphic *Alu* insertion and the Asian origin of Native American populations. *Hum. Biol.* **70**, 23–39.

Ota T. 1993 DISPAN: Genetic distance and phylogenetic analysis, Pennsylvania state University, USA.

Ramaiah A. 2004 Untouchabililty and inter caste relations in rural India: the case of southern Tamil Village. *J. Relig. Cult.* **70**, 1–13.

Sanghvi L. D., Balakrishnan V. and Karve I. 1981 *Biology of the people of Tamil Nadu*, Indian society of Human Genetics, Pune, and The Indian anthropological society, Kolkata.

Singh K. S. 1998 *People of India*, India's communities, national series volume V, Anthropological survey of India, Oxford University press, Delhi.

Stoneking M., Fontius J. J., Clifford S. L. Soodyall H., Arcot S. S., Saha N. *et al.* 1997 *Alu* insertion polymorphisms and human evolutions: evidence for a larger population size in Africa. *Genome Res.* **7**, 1061–1071.

Tamil Nadu Second Backward Class Commission (1989) Socio-educational cum economic survey - state level castewise table, Government of Tamil Nadu.

Tishkoff S. A, Ruano G. and Kidd J. R. 1996 Distribution and frequency of a polymorphic *Alu* insertion at the Plasminogen activator locus in humans. *Hum. Genet.* **97**, 759–764.

Viswanathan H., Edwin D., Cordaux R., Stoneking M., Usha Rani M. V. and Majumder P. P. 2004 Genetic structure and affinities among tribal populations of southern India: a study of 24 autosomal DNA markers. *Ann. Hum. Genet.* **68**, 128–138.

Zischler N., Geisert H., Haeseler A. V. and Paabo S. 1995 A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. *Nature* **378**, 489–492.