RESEARCH NOTE

Genetic study of Dravidian castes of Tamil Nadu

S. KANTHIMATHI, M. VIJAYA and A. RAMESH

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

Introduction

The origin and settlement of Indian people still intrigues scientists studying the impact of past and modern migrations on the genetic diversity and structure of contemporary populations. About 10,000 years ago, proto-Dravidian Neolithic farmers from Afghanistan entered the Indian subcontinent, and were later displaced southwards by a large influx of Indo–European speakers ~ 3500 years ago (Majumder *et al*. 1999). The present study aims to describe the genetic diversity and relationships between the Dravidian caste populations of Tamil Nadu, in an attempt to better understand the contemporary people of this state. We studied nine human-specific indels (insertion/deletion polymorphisms) in DNA samples from 10 Tamil Nadu endogamous groups using phylogenetic and principal component analysis. The genetic affinities of the caste populations of India do not correlate well with socio–cultural rankings.

Indian populations are culturally stratified as tribes and castes. A caste system is a social system in which people are ranked into groups based on heredity, within rigid systems of social stratification. Social classes are defined by a number of endogamous groups often termed as 'jatis', where endogamy is the practice of marrying within a social groups, classes or ethnicities. The contemporary caste populations in India predominantly speak languages that belong to the Indo–European or Dravidian families (Singh 1998). The Dravidian linguistic family includes about 28 languages with 145 million speakers. There are considerable debate about whether Dravidian languages owe their origin to Neolithic peoples of southern India, or were brought into India. In spite of this relative linguistic homogeneity in southern India, cultural barriers due to endogamy have probably been a major reason for genetic diversification among the people of this region (Gadgil *et al*. 1998). Thus, it is of considerable interest to understand the genetic structuring and relationships of southern populations.

*For correspondence. E-mail: kanthi@mail.com.

The present study was carried out in Tamil Nadu, one of southern states of India, with a population of about 62 million people (Census of India 2001). Based on the religion, caste and socio-economic status, over 400 endogamous groups are present in this state, and their size and distribution varies widely. They are known to have received extensive gene flow from different caste and linguistic groups of other regions of India. Thus, the biological status of the present-day groups can be considered as 'immigrants' at varying periods of time. Many of the caste groups have subcastes maintaining endogamous status to some extent (Singh 1998).

Insertion by retroposition of mobile genomic elements such as the *Alu* family is a dynamic type of genetic change in the human genome (Rowold and Herrera 2000). Some human-specific *Alu* elements are considered to be identical by descent (Batzer *et al*. 1995). *Alu* elements are stable genetic markers, and the frequency distribution of polymorphic *Alu* elements 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 improve and greatly facilitate phylogenetic analysis. *Alu* insertion polymorphisms are of great interest for ascertaining the degree of population genetic differentiation (Szmulewicz *et al*. 1999).

We have genotyped 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 base pairs (bp) of a 285-bp *Alu* element at the *CD4* locus (Edwards and Gibbs 1992) for 10 caste populations of Tamil Nadu. To determine the genetic relationships of the caste populations of this state, in general, an additional data set of eight indels from earlier studies (Basu *et al*. 2003) were used in the analysis.

Materials and methods

We have collected a total of 497 unrelated individuals belonging to 10 caste groups of Tamil Nadu, under the guidance of an anthropologist (figure 1 in electronic supplemen-

Keywords. Dravidians; indels; genetic differentiation; Tamil Nadu; caste populations.

tary material at (http://www.ias.ac.in/jgenet/)). These groups are at different levels of social hierarchy (middle and lower), and speak Dravidian languages. The sample size, location of sampling, and anthropological information of the present study populations are furnished in table 1a of electronic supplementary material. Linguistic, historical, demographic, and genetic information about these populations has been described elsewhere (Thurston 1909; Singh 1998).

After obtaining prior informed consent from the individuals, blood samples (5 ml 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 nine autosomal indels. DNA samples were amplified by polymerase chain reaction (PCR) using locus-specific primers for the eight *Alu* elements (*Alu APO, Alu CD4, Alu PV92, Alu TPA25, Alu FXIIIB, Alu ACE, Alu PLAT* and *Alu D1*) and nuclear insertion of mitochondrial DNA segment (*mtNUC*; Zischler *et al*. 1995). The primer sequences and PCR protocols used in this study are given in Majumder *et al*. (1999) and Batzer *et al*. (1995). Amplified PCR products were run on agarose gel, stained with ethidium bromide, and visualized under UV light.

Allele frequencies of the nine loci 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 affinity among the populations and dendrograms were constructed by neighbour-joining (NJ) method with boot strap using DISPAN software (Ota 1993). Principal component analysis was performed using the allelic frequencies observed for the eight indels, and implemented on SPSS 10.1.

Results

Significant departures from Hardy–Weinberg equilibrium expectations were observed for 6 of 90 comparisons. Around five comparisons would be expected to be significant at the 5% level by chance alone, and because none of the significant departures cluster by locus or by population, we consider these to represent normal statistical fluctuations ($P < 0.05$). All the loci showed high levels of polymorphism in the populations studied except *Alu CD4* locus.

Locus and average population heterozygosities computed based on allele frequencies are presented in table 2 of electronic supplementary material. Among the loci, *Alu FXIIIB* showed the highest heterozygosity in the populations (0.455–0.500) and *Alu CD4* the lowest (0.039–0.337). In many cases, the maximum attainable value (0.5) of heterozygosity for a bi-allelic marker is actually attained. The average heterozygosities that reflect within-population heterozygosity are found to be very similar among the study groups.

The analyses of gene diversity, separately for each locus and for all loci taken together, are presented in table 1. The total genomic diversity (H_T) among the populations is observed to be quite high, except for *Alu CD4* locus. It is evident that most of the genomic diversity is due to the contribution of variation within populations (H_S) rather than genetic differences between populations (G_{ST} values). The percentage of genomic diversity attributable to between-population variation, relative to the total genomic diversity, ranges from 1.5% for *Alu PV92* to 5.3% for the *Alu FXIIIB* locus. When all loci are jointly considered, about 3.4% of the total genomic diversity is attributable to between-population variation.

Table 1. Results of gene diversity analysis for individual loci and for all loci considered jointly.

Locus	H_T	H_S	G_{ST}
mtNUC	0.4913	0.4733	0.0367
Alu ACE	0.3206	0.3131	0.0234
Alu APO	0.1475	0.1406	0.0473
Alu FXIIIB	0.3773	0.3572	0.0532
Alu D1	0.3911	0.3731	0.0462
Alu CD4	0.4908	0.4780	0.0261
Alu PLAT	0.4978	0.4772	0.0413
Alu TPA25	0.5000	0.4856	0.0287
Alu PV92	0.4852	0.4778	0.0152
All loci	0.4113	0.3973	0.0340

G_{ST}, between population genomic diversity; H_S, diversity between individuals within populations; H_T, total genomic diversity among the populations.

Neighbour-joining tree is done based on the estimated D_A distances, which were calculated on the basis of allele frequencies of eight loci (excluding *Alu TPA25*, data not available for comparison). To determine the genetic relationships of the caste populations of Tamil Nadu in general, the additional data of eight indels of earlier studied caste populations (Basu *et al*. 2003; Majumder *et al*. 1999, table 1b in electronic supplementary material) were used. The NJ tree of the 26 caste populations of India (16 belongs to Tamil Nadu) is presented in figure 1.

The phylogenetic analysis shows two major clusters, of which one cluster is formed by seven of 10 Dravidian present study groups. The Thevar groups (middle class)—Kallar, Maravar and Agamudayar—formed a separate cluster (Kanthimathi *et al*. 2007). The Gavara Naidu and Reddiyar are the two Telugu-speaking (middle class) migrants from Andhra Pradesh who showed close affinity. Meenavar, a marine fisher-folk (middle class) and Parayan (lower class) are the major outliers.

The second cluster bifurcated into two major clades. The Dravidian upper class groups—Iyer and Iyengar—showed closeness with Indo–Aryans, especially Iyer with West Bengal Brahmin (upper class), in one clade. In the other clade,

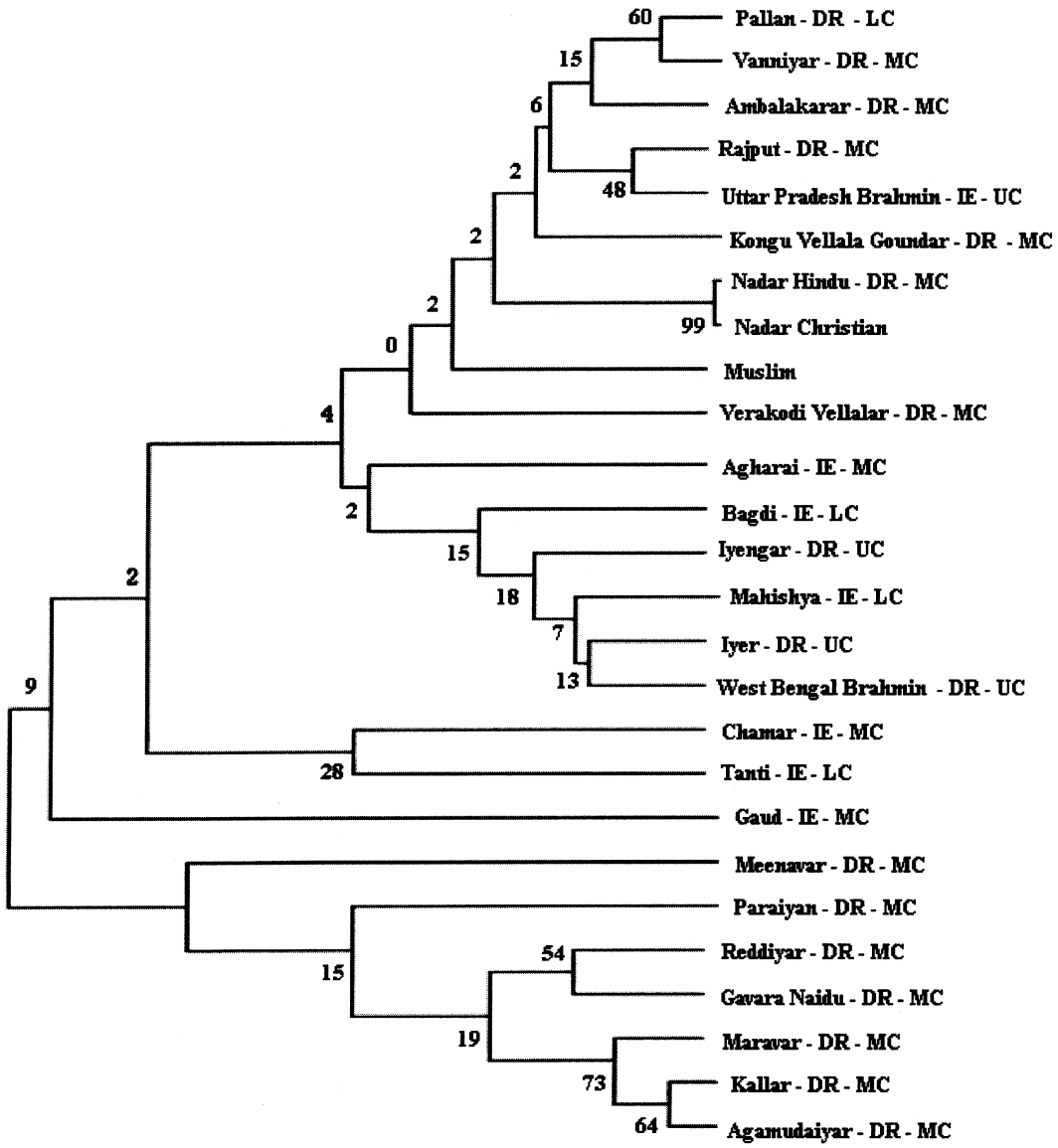

Figure 1. Unrooted neighbour-joining tree depicting genomic affinities of 26 caste populations of India (16 belong to Tamil Nadu caste populations) based on allele frequency data of eight loci. Numbers shown in figures are bootstrap numbers. Linguistical groups: DR, Dravidian; IE, Indo–European; UC, upper class; MC, middle class; LC, lower class.

the groups were mostly Dravidian, except Rajput (middle class), Muslim, and Uttar Pradesh Brahmin (upper class). The two Nadar groups (middle class)—Nadar Hindu and Nadar Christian—belonging to different religions are closely related. This may be due to relatively recent conversion to Christianity within the Nadar group. The first reported conversion occurred during early 19th century (Sanghvi and Balakrishnan 1981).

Principal components analysis of the allele frequencies at the eight polymorphic *Alu* insertion loci was also performed (figure 2). The two principal components explain about 89 per cent of the allele variance, which provide the most information for a two-dimensional depiction of population relationships. The grouping of populations was almost consistent with the clustering pattern of the NJ tree (figure 1).

Discussion

The average heterozygosity values of the present study populations were comparable with other previously studied caste groups of Tamil Nadu (Basu *et al*. 2003). Interestingly, the average heterozygosity levels were higher, although not always significant, than in most global populations studied, with the exception of African populations (Stoneking *et al*. 1997). Thus, the DNA markers attest that the caste

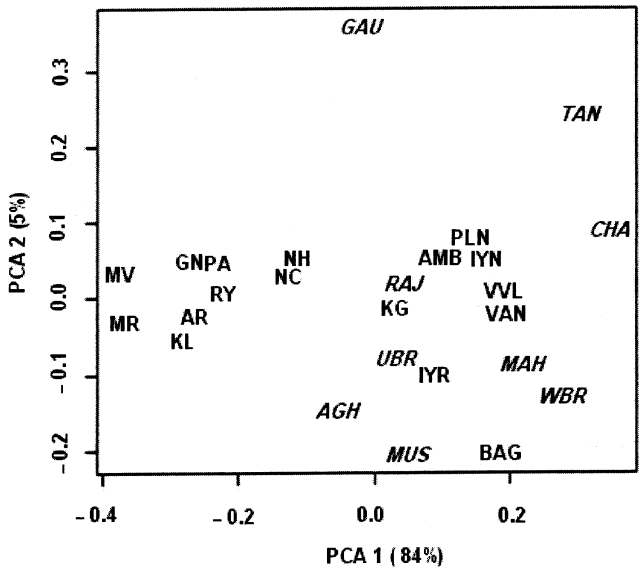

Figure 2. Principal component analysis of 26 caste populations of India (16 belong to Tamil Nadu caste populations) based on allele frequency data of eight loci. Dravidians: KG, Kongu Vellala Gounder; KL, Kallar; MR, Maravar; PA, Parayan; RY, Reddiyar; GN, Gavara Naidu; AR, Agamudaiyar; MV, Meenavar; NH, Nadar Hindu; NC, Nadar Christian; PLN, Pallan; AMB, Ambalakarar; VAN, Vanniyar. Indo-Europeans: AGG, Agharia; BAG, Bagdi; UBR, Brahmin (UP); WBR, Brahmin (WB); CHA, Chamer; GAU, Gaud; MAH, Mahishya; MUS, Muslim; RAJ, Rajput; TAN, Tanti.

populations of the state exhibit high levels of genomic diversity, which is in concordance with the earlier study of Tamil Nadu tribal populations (Vishwanathan *et al*. 2004).

There is significantly greater interindividual variation within each caste population (G_{ST} = 3.4%) than between them. At the continent level, the observed G_{ST} values are much lower than the Africans (8.8%) and Southeast Asians (5.8%), but higher than the Europeans (1.1%), and populations from Australia and New Guinea (0.1%; Stoneking *et al*. 1997). A significantly greater interindividual variation observed within each study population and moderate population differentiation probably indicates relative genetic closeness of these populations with much more diverse groups of the Indian subcontinent.

Thus, the caste populations of Tamil Nadu are characterized by a high levels of heterozygosity and almost similar allele frequency profiles. This suggests that these populations might have a common ancestry or probably experienced very high gene flow during the period of their coexistence. The above finding is substantiated by moderate genetic differentiation irrespective of the social diversity of the caste groups and their varying migration periods.

The phylogenetic analysis of the caste populations of India showed that Dravidians show closer affinities among themselves than with Indo–Aryans, irrespective of their position in the social hierarchy. This is in concordance with the earlier findings (Basu *et al*. 2003). The most recent migrants among Dravidians of Tamil Nadu (upper class Iyer and Iyengar) showed close affinity with Indo–Aryans (West Bengal Brahmin, upper class; Mahisya, middle class; and Bagdi, lower class) than the early migrants into Tamil Nadu. These findings support the earlier evidence that the Dravidians were possibly widespread throughout India before the arrival of the Indo–Aryans and may have retreated to southern India to avoid linguistic dominance, after an initial period of admixture and adoption of the caste system (Basu *et al*. 2003). The present study shows that these contemporary caste populations of Tamil Nadu are confounded by assimilation of subsequent immigrants in varying degrees of admixture before the caste system became too rigid.

Acknowledgements

We are grateful to the donors of the samples. We are also grateful to the Department of Biotechnology, Government of India, for their financial support.

References

Basu A., Mukherjee N., Roy S., Sengupta S., Banerjee S., Chakraborty M. *et al*. 2003 Ethnic India: A genomic view, with special reference to peopling and structure. *Genomic 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. *J. Mol. Bio*. **247**, 418–427.

Census of India 2001 Provisional population total paper - 1 of 2001, series - 34. Directorate of Census Operations, Government of Tamil Nadu.

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

Gadgil M., Joshi N. V., Manoharan P., Patil S. and Prasad U. V. S. 1998 Peopling of India, In *The Indian human heritage* (ed. D. Balasubramanian and N. A. Rao), pp. 100–129. Universities Press, Hyderabad.

Kanthimathi S., Vijaya M., Srikumari S., Reddy G., Majumder P. P. and Ramesh A. 2007 A genetic structure of the early immigrants (Mukkalathor) of Tamil Nadu as inferred from autosomal loci. *Int. J. Hum. Genet*. **7**, 167–173.

Majumder P. P., Roy B., Banerjee S., Chakaraborty M., Dey B., Mukherjee N. *et al*. 1999 Human specific insertion/deletion polymorphism in Indian populations and their possible evolutionary implications. *Eur. J. Hum. Genet*. **7**, 435–446.

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*. **6**, 1215.

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 State University, PA, USA.

Renfrew C. 1987 *Archaeology and language: the puzzle of Indo-European origins*. Jonathan Cape, London.

Rowold D. J. and Herrera R. J. 2000 *Alu* elements and the human genome. *Genetics* **108**, 57–72.

Sanghvi L. D. and Balakrishnan V. 1981 Ethnological notes on the populations investigated. In *Biology of the people of Tamil Nadu* (ed. L. D. Sanghvi, V. Balakrishnan and I. Karve), pp. 10–25. The 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, New Delhi.

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

Szmulewicz M. N., Andino L. M., Reategui E. P., Woolley-Barker T., Jolly C. J., Disotell T. R. and Herrera R. J. 1999 An *Alu* insertion polymorphism in a baboon hybrid zone. *Am. J. Phys. Anthropol*. **109**, 1–8.

Thurston E. 1909 *Castes and tribes of South India*, Government Press, Chennai.

Vishwanathan 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.

Received 23 September 2007, in revised form 5 February 2008; accepted 5 February 2008
Published on the Web: 20 June 2008