Genetic structure and affinities among tribal populations of southern India: a study of 24 autosomal DNA markers

H. Vishwanathan[1,*], E. Deepa[1], R. Cordaux[2], M. Stoneking[2], M. V. Usha Rani[1] and P. P. Majumder[3]

[1]*Division of Genetics, Department of Environmental Sciences, Bharathiar University, Coimbatore − 641 046, India*
[2]*Max Planck Institute for Evolutionary Anthropology, Department of Evolutionary Genetics, 04103 Leipzig, Germany*
[3]*Anthropology and Human Genetics Unit, Indian Statistical Institute, Kolkata − 700 035, India*

Summary

We describe the genetic structure and affinities of five Dravidian-speaking tribal populations inhabiting the Nilgiri hills of Tamil Nadu, in south India, using 24 autosomal DNA markers. Our goals were: (i) to examine what evolutionary forces have most significantly impacted south Indian tribal genetic variation, and (ii) to test whether the phenotypic similarities of some south Indian tribal groups to Africans represent a signature of close relationship to Africans or are due to convergence. All loci were polymorphic and average heterozygosities were substantial (range: 0.347–0.423). Genetic differentiation was high ($G_{st} = 6.7\%$) and genetic distances were not significantly correlated with geographic distances. Genetic drift therefore probably played a significant role in shaping the patterns of genetic variation observed in southern Indian tribal populations. Otherwise, analyses of population relationships showed that Indian populations are closely related to one another, regardless of phenotypic characteristics, and do not show particular affinities to Africans. We conclude that the phenotypic similarities of some Indian groups to Africans do not reflect a close relationship between these groups, but are better explained by convergence.

Introduction

Contemporary ethnic populations of India are highly variable, both biologically and culturally (Majumder, 1998). The broadest division of Indian populations distinguishes tribal from non-tribal groups. The definition of tribe is somewhat ambiguous, but generally refers to the endogamous populations that are considered aboriginal, inhabiting the Indian subcontinent before the immigration of pastoral nomads from central Asia some 3,500 years ago (Cavalli-Sforza *et al.* 1994). The tribal groups constitute about 8% of the total Indian population and they "may represent relic populations of unknown origin but potentially of great ge-

netic interest" (Cavalli-Sforza *et al.* 1994). The origins and migrational histories of the tribal populations of the Indian subcontinent are not clearly understood. It has been argued that Africa may have made some direct genetic contribution to India, since some tribal populations in southern India possess phenotypic similarities with Africans, the so-called "Negrito" physical characteristics (Maloney, 1974; Saha *et al.* 1974; Roychoudhury, 1982; Chandler, 1988; Majumder, 1998). It has also been suggested that at one time a "Negrito element" was widespread throughout India and was eventually forced into a more restricted location in south India (Majumder & Mukherjee, 1993).

Analyses of mitochondrial DNA (mtDNA) and Y-chromosomal variation in south India suggest that contemporary genetic variation may derive from the original Indian gene pool (Roychoudhury *et al.* 2001; Cordaux *et al.* 2003; Kivisild *et al.* 2003), although the

*Corresponding author: Dr. H. Vishwanathan, Division of Genetics, Department of Environmental Sciences, Bharathiar University, Coimbatore 641 046, TN, India. Phone: +91-422-2422222 ext.395. Fax: +91-422-2422387. E-mail: bharathvishwa@yahoo.co.uk

time depth for the arrival of the first settlers of the Indian subcontinent remains a matter of debate (e.g. Cordaux & Stoneking, 2003; Endicott *et al.* 2003b). The presence of mtDNA haplogroup M in both east Africa and India has been interpreted as supporting an ancient east Africa-to-India migration (Quintana-Murci *et al.* 1999), although this conclusion has been questioned (Roychoudhury *et al.* 2001). Moreover, comparisons of south Indian tribal groups with other world populations, based on mtDNA data, do not reveal close relationships to Africans (Cordaux *et al.* 2003; Kivisild *et al.* 2003). This conclusion is also supported by Y-chromosome genetic markers (Kivisild *et al.* 2003; Cordaux *et al.* 2004). Therefore, it is possible that the African-like morphological features of some south Indian tribal groups result from convergence rather than shared ancestry with Africans.

However, mtDNA and the Y-chromosome each represent single haploid loci and are more prone to stochastic processes than are autosomal bi-parentally inherited markers. Here, we report the analysis of 24 autosomal markers (including 7 insertion-deletion polymorphisms and 17 restriction site polymorphisms or RSPs) in five indigenous tribal populations from the Nilgiri Hills of Tamil Nadu, in southern India. The aims of this study were: (i) with regard to the relationships of these populations, to test whether the "Negrito" features of some contemporary south Indian tribal groups represent a signature of close relationship to Africans or are due to convergence, and (ii) with regard to the genetic structure of these populations, to examine what evolutionary forces have most significantly impacted on the genetic variation in present-day south Indian tribes.

Materials and Methods

Population Samples and Autosomal Markers

Blood samples (5-10 ml by venipuncture) were drawn from 250 unrelated adult volunteers from five endogamous Dravidian-speaking tribal populations of southern India with prior informed consent. The tribal groups are confined to hilly tracts and valleys of the Nilgiri region (Figure 1), located 20–100 km from each other (Table 1), and consist of Irula (n = 50), Kurumba (n = 54), Badaga

(n = 51), Kota (n = 45) and Toda (n = 50). Of these, the Irula and Kurumba possess "Negrito" morphological features (for a description, see Majumder, 1998). These groups are characterized by small census sizes, in particular Toda (census size: ∼1,300 individuals), Kota (∼2,000), Kurumba (∼5,000) and Irula (∼9,000), whereas Badaga constitute a larger group (∼150,000). Additional linguistic, historical, demographic and genetic information about these populations have been reported elsewhere (Thurston, 1909; Saha *et al.* 1976; Breeks, 1983; Singh, 1994; Roychoudhury *et al.* 2000).

High molecular weight DNA was isolated from the blood samples by the salting out procedure (Miller *et al.* 1988) and was suspended in 10mM Tris and 0.1mM EDTA for genotyping. All the polymorphic loci studied were genotyped by amplifying DNA in a standard 30-cycle three-step PCR. Appropriate annealing temperatures and additives were optimized for each system. PCR products of the RSPs were digested with 5 U of the appropriate restriction enzymes in the respective buffers for 2–4 hours at the appropriate temperatures. The names and Genome Database (GDB) accession numbers of the 9 RSP loci are: ESR1 (GDB: 185229), NAT (GDB: 187676), PSCR (GDB: 182305), T2 (GDB: 196856), LPL (GDB: 285016), ALB (GDB: 178648), CYP1A1 (GDB: 120604), HoxB4 (GDB: 120663) *Msp1*, ADH2 (GDB: 119651) *Rsa1*. The chromosomal locations, primer sequences and PCR conditions for each RSP locus studied are provided at GDB and are listed on the Eccles Institute of Human Genetics website (http:// www. genetics.utah.edu/∼swatkins/pub/RSP_links.html). The methodologies of the haplotype loci DRD2 (based on sites Taq1 'A', Taq1 'B' and Taq1 'D'), β-globin (based on sites HB7, HB8 and HB9) and ALAD (based on sites RSA1 and MSP1) and RSPs are also described elsewhere (Kidd *et al.* 1998; Majumder *et al.* 1999a; Jorde *et al.* 1995; Mukherjee *et al.* 2000; Wetmur *et al.* 1991). The insertion-deletion polymorphisms include 5 *Alu* elements (*Alu*APO, *Alu*ACE, *Alu*PLAT, *Alu*PV92, *Alu*FXIIIB), one deletion within the *Alu* element CD4 (CD4del) and one nuclear insertion of a mitochondrial DNA segment (mtNUC; Zischler *et al.* 1995). The protocols for these markers have been described elsewhere (Stoneking *et al.* 1997; Majumder *et al.* 1999b; Watkins *et al.* 2001). After PCR, the

Figure 1 Map of India pinpointing the Nilgiri Hills of Tamil Nadu, in southern India.

samples were subjected to electrophoresis at 125 V for 1 hour. Ethidium bromide stained gels were visualized by UV and were documented.

Data Analysis

Allele frequencies were calculated by direct counting at each locus separately for each population. Heterozygosi-ties at individual loci and the overall average heterozy-gosity were calculated using the estimated allele frequen-cies for each population. Hardy–Weinberg equilibrium was tested using a χ^2 goodness of fit test, with Bonfer-roni's correction for multiple comparisons. To assess the extent of gene differentiation among the studied groups, Nei's (1973) measure of gene diversity was calculated

Table 1 Matrices of pairwise distances among five southern Indian tribal populations

	Badaga	Irula	Kota	Kurumba	Toda
Badaga	–	90	20	90	45
Irula	0.032	–	100	20	100
Kota	0.043	0.041	–	100	70
Kurumba	0.044	0.030	0.063	–	100
Toda	0.060	0.075	0.058	0.106	–

Above the diagonal are approximate geographic distances (km) separating populations. Below the diagonal are Nei's (1972) standard genetic distances based on 24 autosomal markers.

separately for each locus and for all loci considered jointly. A contingency χ^2 analysis (Workman & Niswander, 1970) was used to test for heterogeneity in the populations, which reflects differences in allele frequencies. Maximum likelihood estimates of the haplotype frequencies were calculated for the multisite marker typing data, using the program HAPLOFREQ (Majumder & Majumder, 2000). To assess genetic relationships among the populations, multidimensional scaling (MDS) analyses were performed by means of STATISTICA (Statsoft Inc., Tulsa, OK, USA), based on Nei's (1972) standard genetic distance calculated with PHYLIP version 3.5c (Felsenstein, 1993). In addition, dendrograms were constructed using the neighbor-joining (NJ) method, based on Nei's (1972) standard genetic distance.

Results

Genetic Diversity and Population Differentiation

All of the loci were polymorphic in all of the populations (Table 2), with the exception of *Alu*APO in the Toda and CD4del in the Kota and Kurumba. All the RSPs and haplotype loci studied were also polymorphic in all populations. None of 117 χ^2 tests for goodness of fit to Hardy-Weinberg equilibrium showed significant departure after Bonferroni correction for multiple tests was applied. The average heterozygosity for each locus was substantial, with several values approaching the theoretical maximum heterozygosity of 0.5 for a biallelic locus, with the notable exception of the CD4del locus which showed low and consistently minimum heterozygosity in all the populations (results not shown). Although high

Table 2 Allele frequencies for 24 autosomal markers and average heterozygosity (*h*) estimates in five southern Indian tribal populations

Markers	Populations				
	Badaga	Irula	Kota	Kurumba	Toda
mtNUC	0.625	0.574	0.567	0.538	0.204
*Alu*PV92	0.436	0.449	0.300	0.713	0.255
*Alu*FXIIIB	0.461	0.640	0.878	0.694	0.806
*Alu*APO	0.784	0.570	0.767	0.583	1.000
*Alu*ACE	0.460	0.750	0.622	0.806	0.469
*Alu*PLAT	0.551	0.550	0.659	0.704	0.406
CD4del	0.020	0.040	0.000	0.000	0.050
ESR	0.684	0.673	0.333	0.759	0.396
NAT	0.708	0.760	0.811	0.889	0.688
PSCR	0.239	0.330	0.244	0.352	0.207
T2	0.590	0.680	0.767	0.759	0.888
LPL	0.716	0.540	0.789	0.731	0.436
ALB	0.552	0.398	0.744	0.377	0.533
CYP 1A	0.322	0.612	0.422	0.417	0.184
HOX B4	0.470	0.510	0.556	0.404	0.528
ADH2	0.544	0.359	0.289	0.314	0.553
ALAD/RSA	0.310	0.163	0.178	0.481	0.271
ALAD/MSP	0.061	0.224	0.233	0.222	0.109
DRD2/Taq1 'A'	0.590	0.633	0.489	0.537	0.798
DRD2/Taq1 'B'	0.657	0.612	0.767	0.667	0.909
DRD2/Taq1 'D'	0.652	0.570	0.733	0.500	0.439
β-globin/HB7	0.340	0.220	0.311	0.426	0.245
β-globin/HB8	0.480	0.540	0.444	0.565	0.281
β-globin/HB9	0.814	0.900	0.989	0.704	0.959
Average *h*	0.422	0.423	0.372	0.415	0.347

The frequency indicated for each bi-allelic marker is that of the presence of the insert for insertion-deletion markers except CD4del; presence of the deletion for CD4del; presence of the restriction site for RSPs.

heterozygosity was expected because these loci were ascertained on the basis that they were known to be polymorphic, an analysis of the distribution of allele frequencies at loci ascertained in a similar manner indicates that ascertainment bias alone does not completely account for the observed frequency spectrum; the distributions also contain information on the demographic history of human populations (Sherry *et al.* 1997). Overall, the average heterozygosity for each population ranged from 0.347 (in Toda) to 0.423 (in Irula), the lowest values being found in the Toda and Kota (Table 2).

To determine the amount of genetic differentiation among populations, G_{st} values (a measure of the interpopulation variability) for each polymorphic locus were determined. The results are presented in Table 3,

Table 3 Gene diversity analysis for individual loci and for all loci considered jointly

Loci	H_T	H_S	G_{ST}
*Alu*ACE	0.471	0.431	0.085
*Alu*FXIIIB	0.423	0.381	0.098
*Alu*APO	0.384	0.335	0.129
CD4del	0.043	0.042	0.019
*Alu*PLAT	0.489	0.468	0.044
*Alu*PV92	0.490	0.439	0.104
mtNUC	0.499	0.454	0.092
ESR	0.490	0.432	0.119
NAT	0.353	0.342	0.030
PSCR	0.398	0.392	0.016
T2	0.388	0.368	0.051
LPL	0.459	0.424	0.076
ALB	0.499	0.464	0.069
CYP 1A	0.476	0.437	0.082
HOX B4	0.499	0.494	0.011
ADH2	0.484	0.459	0.054
ALAD	0.592	0.558	0.058
DRD2	0.753	0.702	0.068
β-globin	0.674	0.649	0.036
All loci	0.466	0.435	0.067

separately for each locus and also for all loci taken together. Except for CD4del, the total genomic diversity (H_T) among the subpopulations was quite high. However, most of the genetic diversity is attributable to diversity between individuals within populations (H_S). The percentage of the total genetic diversity attributable to differences between populations ranged from 1.1% for HOXB4 to 12.9% for *Alu*APO. When all loci are jointly considered, 6.7% of the total genetic diversity is attributable to variation between populations. Tests of significance for heterogeneity of populations also showed significant values for 85 out of 160 comparisons involving insertion/deletion loci and RSPs (data not shown). The three haplotype loci showed significant values in all the comparisons, with the exception of the Badaga/Toda and Irula/Kota for the ALAD locus. Thus, there are substantial differences among populations with respect to their allele frequencies.

Population Relationships

The high levels of differentiation observed among the groups studied could be explained by: (i) a common ancestry, but genetic drift has altered ancestral gene frequencies, or (ii) independent origins. To distinguish be-

Table 4 Haplotype frequency estimates in five southern Indian tribal populations

Haplotypes	Populations				
	Badaga	Irula	Kota	Kurumba	Toda
a. DRD2					
B2D2A2	0.313	0.162	0.441	0.145	0.300
B2D1A2	0.273	0.429	0.048	0.381	0.495
B1D2A2	0.000	0.042	0.000	0.000	0.000
B1D1A2	0.012	0.000	0.000	0.011	0.000
B2D2A1	0.078	0.021	0.059	0.080	0.033
B2D1A1	0.000	0.000	0.219	0.060	0.082
B1D2A1	0.262	0.346	0.233	0.275	0.090
B1D1A1	0.063	0.000	0.000	0.047	0.000
b. β-globin					
$(+ + +)$	0.203	0.162	0.311	0.154	0.147
$(+ + -)$	0.137	0.058	0.000	0.272	0.014
$(+ - +)$	0.000	0.000	0.000	0.000	0.088
$(+ - -)$	0.000	0.000	0.000	0.000	0.000
$(- + +)$	0.123	0.290	0.122	0.133	0.120
$(- + -)$	0.017	0.030	0.011	0.006	0.000
$(- - +)$	0.494	0.447	0.556	0.417	0.603
$(- - -)$	0.026	0.012	0.000	0.018	0.028
c. ALAD					
$(+ +)$	0.000	0.005	0.000	0.000	0.000
$(+ -)$	0.061	0.219	0.233	0.222	0.117
$(- +)$	0.316	0.158	0.178	0.481	0.266
$(- -)$	0.622	0.617	0.589	0.296	0.617

Haplotypes are listed as B1, D1 and A1 alleles for the site-absent state for the RFLP sites while B2, D2 and A2 are the site-present alleles for the DRD2 loci. A polymorphic restriction site is denoted as being present in a particular haplotype by a $(+)$ symbol and absent by a $(-)$ symbol for the β-globin and ALAD loci.

tween these two hypotheses, we first analyzed the distribution of haplotypes at three haplotype loci in the five studied tribal groups. With respect to the DRD2 loci (Table 4a; Vishwanathan *et al.* 2003), four out of eight possible haplotypes (B2D2A2, B2D1A2, B2D2A1 and B1D2A1) were shared by all five populations and accounted for 78–96% of the haplotypes in each group. Except B1D2A2 (which was restricted to the Irula), none of the haplotypes was specific to "Negrito" or "non-Negrito" groups. In the β-globin gene (Table 4b), the $(- - +)$ haplotype was modal in all populations. In addition, except for the $(+ - +)$ haplotype (which was restricted to the Toda), none of the haplotypes was specific to "Negrito" or "non-Negrito" groups. Otherwise, the Hb-S allele occurs primarily on the $(- + +)$ haplotype background (Majumder *et al.* 1999a) which

was present in all the study populations with a frequency ranging from 0.12 to 0.29. Most Hb-A alleles occur on the $(+ \;\; - \;\; -)$ haplotype background which was not seen in any of the studied populations. Out of the four possible haplotypes at the ALAD locus (Table 4c), three haplotypes $(+ \; -)$, $(- \; +)$, $(- \; -)$ were shared by all the populations, whilst the $(+ \; +)$ haplotype was absent or virtually absent in all the study groups. The haplotype $(- \; -)$ was most prevalent in all groups except the Kurumba. Altogether, the extensive sharing of haplotypes by all five groups is more consistent with a scenario of common ancestry than one of independent origins.

To further distinguish between these two hypotheses, the relationships among the study populations were assessed by calculating Nei's (1972) standard genetic distances between populations, based on the 24 markers (Table 1). The distances ranged from 0.030 (between Irula and Kurumba) to 0.106 (between Kurumba and Toda). The Toda showed large genetic distances with all other groups (above 0.055), whereas all but one pairwise distance values not involving the Toda were less than 0.045. A Mantel test indicated that geographic distances (Table 1) are not significantly correlated with the genetic distances among the five study groups ($r^2 =$ 0.435; $p = 0.104$). This lack of significant correlation supports the idea of enhanced genetic drift in these small and isolated populations.

The Nilgiri Hills' tribal Indians were also compared to 11 other tribal, and 10 non-tribal, populations of India, using the data from seven insertion/deletion loci presented by Majumder *et al.* (1999b), Mukherjee *et al.* (2000) and Veerraju *et al.* (2001) that are also analyzed in the present study (i.e. mtNUC, *Alu*PV92, *Alu*FXIIIB, *Alu*APO, *Alu*ACE, *Alu*Plat and CD4del). The MDS plot of the 26 Indian populations is presented in Figure 2. Overall, the tribal groups tend to fall on the left-hand side of the plot, whereas the non-tribal groups tend to fall on the right-hand side. Regarding the study groups, the "Negrito" groups (i.e. Irula and Kurumba) cluster together with other Indian tribal groups, as do the Kota. By contrast, the Toda and Badaga are well separated from the other tribal populations (Figure 2): the Toda are isolated from all other groups whereas the Badaga show closer affinities to non-tribal groups. The latter observation could indicate common ancestry of the Badaga with non-tribal groups, or could represent an artifact generated by extensive genetic drift occurring in the Badaga. Noticeably, in an NJ tree, both the Toda and Badaga were characterized by long branches (not shown), consistent with genetic drift occurring in these populations.

In order to compare the global relationships among populations, we used the available *Alu* insertion data from Stoneking *et al.* (1997) and Majumder *et al.*
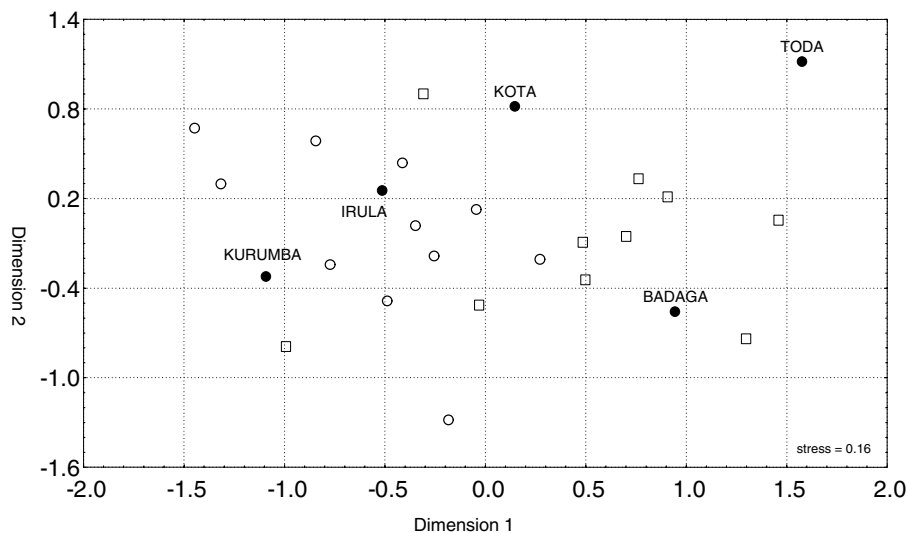

Figure 2 Multidimensional scaling plot depicting the genetic relationships of 26 Indian populations based on seven insertion/deletion loci (mtNUC, *Alu*PV92, *Alu*FXIIIB, *Alu*APO, *Alu*ACE, *Alu*Plat and CD4del). Filled circles represent Nilgiri Hills tribal populations; open circles represent other tribal groups; open squares represent non-tribal groups.
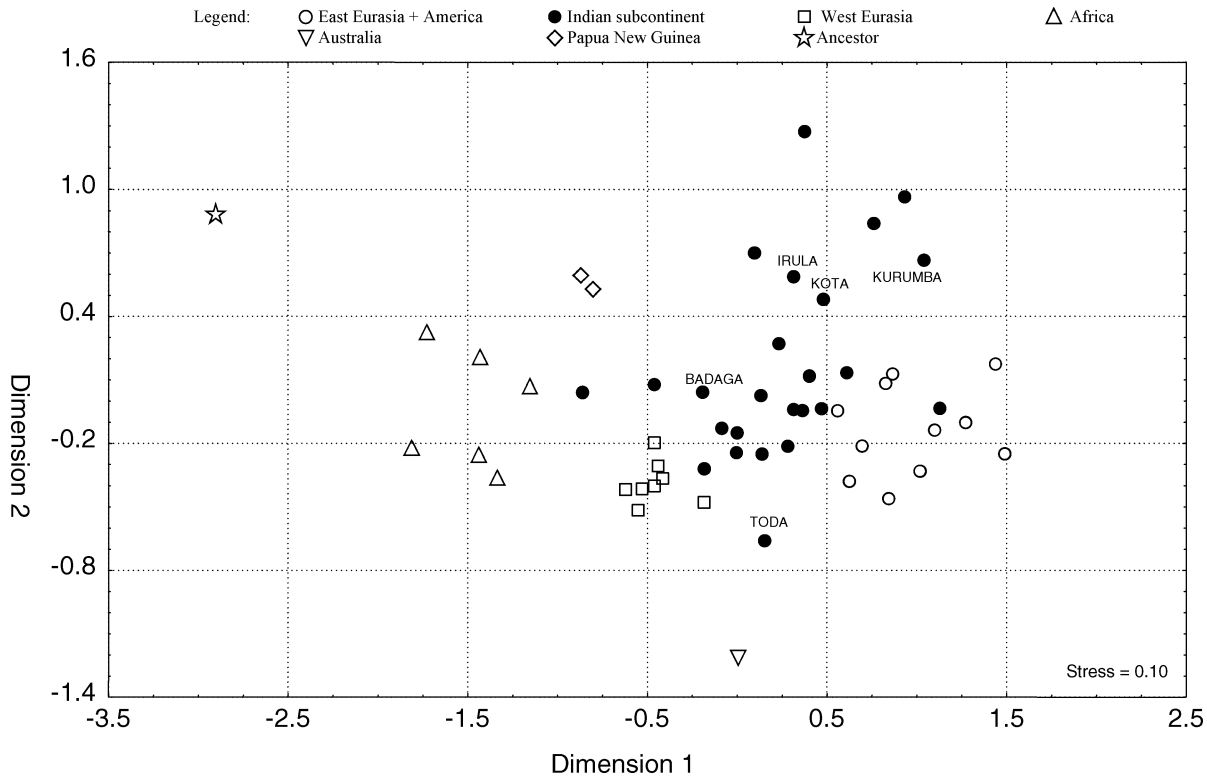
Figure 3 Multidimensional scaling plot depicting the genetic relationships of 54 world populations, using five *Alu* insertions loci (*Alu*ACE, *Alu*FXIIIB, *Alu*APO, *Alu*PV92, *Alu*PLAT). A hypothetical ancestral population is shown, in which the frequency of the *Alu* element at each locus is set to 0.

(1999b) that are common to our dataset (i.e. *Alu*ACE, *Alu*FXIIIB, *Alu*APO, *Alu*PV92, *Alu*PLAT). The overall pattern of the MDS analysis (Figure 3) suggests that Indian populations in general occupy an intermediate position in between west and east Eurasians, as found previously (Majumder *et al.* 1999b). The five Nilgiri Hills' populations do not cluster together, but fall close to other Indian groups. In particular, the Indian "Negrito" populations do not show any particular ties to African populations and are clearly more closely related to other Indian groups than to African groups.

Discussion

It is generally thought that Indian tribal populations are descendants of the original inhabitants of India. Their morphological, social and genetic features, maintained in highly endogamous groups, provide a unique opportunity for examining human evolution and population histories. However, most studies of Indian pop-

ulations using DNA markers have included at most a few tribal groups. Likewise, analyses of Indian populations using a large number of highly varied and selectively neutral markers distributed throughout the genome have primarily focused on non-tribal groups (Bamshad *et al.* 2001, 2003; Watkins *et al.* 2001). Thus, the present investigation was conducted with the goal of analyzing the extent of genetic variation at a number of polymorphic autosomal loci from samples of diverse tribal populations from southern India, with a particular focus on the origins of particular groups that show phenotypic similarities to Africans (i.e. "Negrito" characteristics).

Populations with "Negrito" features have been reported in southern Asia, southeast Asia and island southeast Asia, leading to the suggestion that they might represent the signature of an ancient migration from Africa (Cavalli-Sforza *et al.* 1994). An alternative explanation for these characteristics is phenotypic convergence. If the former hypothesis is correct, one would expect the Indian "Negrito" groups to show closer genetic

affinities to African groups than to other "non-Negrito" Indian groups. By contrast, the reverse would be expected under a scenario of convergence. The comparison of autosomal markers in "Negrito" and "non-Negrito" populations from southern India supports the latter hypothesis. Indeed, our results comparing Indian populations to worldwide populations using *Alu* insertion markers convincingly show that all Nilgiri Hills' tribal groups (regardless of their phenotypic characteristics) are more closely related to other Indian groups than they are to African groups (Figure 3). Interestingly, Endicott *et al*. (2003a) also suggested that the "Negrito" features of the Andaman Islanders (in southeast Asia) are due to convergence rather than to common ancestry with Africans. These results emphasize the limits of morphological classifications for inferring population relationships.

Therefore, our results suggest that southern Indian tribal populations share a common ancestry, although they are morphologically diverse. What are the genetic characteristics of these populations? And what evolutionary forces generated the observed patterns? Polymorphic *Alu* insertions showed high levels of polymorphism in the tribal populations of south India, as found previously for other groups from India (Majumder *et al*. 1999b; Mukherjee *et al*. 2000; Veerraju *et al*. 2001) and elsewhere (e.g. Batzer *et al*. 1994, 1996; Stoneking *et al*. 1997). Similarly, the RSPs and haplotype loci also showed high levels of polymorphism in the studied groups, in good agreement with earlier reports on global populations (Jorde *et al*. 1995; Kidd *et al*. 1998). In the present study, the estimated levels of average heterozygosities were consistently high in all the populations. The heterozygosity levels were similar to those of other Indian populations using nuclear DNA markers (Majumder *et al*. 1999b; Mukherjee *et al*. 2000; Veerraju *et al*. 2001). Interestingly, the average heterozygosity levels were higher than in other global populations studied (Europe and America), with the exception of African populations (Stoneking *et al*. 1997; Novick *et al*. 1998). Thus, the autosomal DNA markers attest that the study groups exhibit high levels of genetic diversity.

The extent of genetic differentiation based on 24 polymorphic markers for the five southern Indian tribal populations ($G_{st} = 6.7\%$) is higher than that observed in other parts of India (Mukherjee *et al*. 2000; Veerraju *et al*. 2001), but smaller than continental-level estimates based on autosomal RSPs and microsatellites (Bowcock *et al*. 1991; Deka *et al*. 1995; Barbujani *et al*. 1997; Stoneking *et al*. 1997; Novick *et al*. 1998; Jorde *et al*. 2000). Watkins *et al*. (2001) reported a G_{st} value of 2.4% for 12 Indian populations using *Alu* insertion polymorphisms, which is about one third of the G_{st} value estimated in the present study. Thus a substantial interpopulation variability was observed in the study populations. Various analyses based on genetic distances among populations corroborated this finding. Furthermore, the latter analyses also suggested that the process of genetic differentiation of the south Indian tribal groups has been accentuated by genetic drift. The likely key-role of genetic drift in shaping genetic variation in south India is further supported by dramatic allele frequency differences observed in the populations. In addition, some groups such as the Toda showed large genetic distances to their other tribal neighbors, as well as smaller average heterozygosity levels. Through the passage of time, drift must have played an important role in the genetic differentiation of these small and isolated populations (i.e. four out of the five study groups have census sizes less than 10,000).

Under drift conditions, one would expect not only high differentiation among populations, but also reduced diversity. However, we have presented evidence that, with respect to autosomal DNA polymorphisms, the south Indian populations show high levels of heterozygosity. This observation may be explained by population history, for example by a high inflow of genes into the study populations (resulting in high heterozygosities), but different populations having received alleles from different sources (resulting in high levels of genetic differentiation). However, there is no evidence from the anthropological literature to support the hypothesis that the different study populations have had an inflow of genes from different external sources, and our genetic analyses suggest a common ancestry for all Indian groups. Another possible scenario involves an early inflow of genes into a population, followed by a rapid expansion of this population (resulting in high heterozygosities), and subsequent splits of this population into largely isolated (endogamous) populations (resulting in high levels of genetic differentiation). A third

possibility is that the high diversity suggested by autosomal markers may reflect a lack of sensitivity and/or ascertainment bias of these bi-allelic markers, since they were selected because they are known to be polymorphic in the Indian subcontinent. Mitochondrial DNA analyses in other south Indian tribal groups support this explanation, as both high differentiation between groups and low diversity within groups is observed (Cordaux *et al.* 2003).

Thus, both mtDNA and autosomal evidence point to a high genetic differentiation of south Indian tribal groups, consistent with the occurrence of drift in these populations. However, it remains to be determined whether the effects of genetic drift result from a recent and strong bottleneck (possibly as a consequence of the Indo-European migration to India some 3,500 years ago; Cordaux *et al.* 2003) or from the long-lasting maintenance of a small and constant population size. It is conceivable that both hypotheses are correct, since mtDNA sequence data are more likely to capture recent demographic events, whereas autosomal markers, because of their fourfold larger effective population size, are less sensitive to recent events and more likely to reflect older or long-term demographic events.

In conclusion, the present study suggests that the tribal groups of southern India share a common ancestry, regardless of phenotypic characteristics, and are more closely related to other Indian groups than to African groups. Based on 24 autosomal loci, they appear to show high levels of genetic diversity and genetic differentiation. Several lines of evidence suggest that genetic drift has been the major evolutionary force to shape genetic variation in these populations. This represents an important feature of tribal populations of south India, which has to be taken into account in any attempt to reconstruct the history of these populations.

Acknowledgments

References

Bamshad, M., Kivisild, T., Watkins, W. S., Dixon, M. E., Ricker, C. E., Rao, B. B., Naidu, J. M., Prasad, B. V. R., Reddy, P. G., Rasanayagam, A., Papiha, S. S., Villems, R., Redd, A. J., Hammer, M. F., Nguyen, S. V., Carroll, M. L., Batzer, M. A. & Jorde, L. B. (2001) Genetic evidence on the origins of Indian caste populations. *Genome Res* **11**, 994–1004.

Bamshad, M. J., Wooding, S., Watkins, W. S., Ostler, C., Batzer, M. A. & Jorde, L. B. (2003) Population genetic structure and human classification *Am J Hum Genet* **72**, 578–89.

Barbujani, G., Magagni, A., Minch, E. & Cavalli-Sforza, L. L. (1997) An apportionment of human DNA diversity. *Proc Natl Acad Sci USA* **94**, 4516–4519.

Batzer, M. A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D. H., Shaikh, T. H., Novick, G. E., Ioannou, P. A., Scheer, W. D., Herrera, R. J. & Deininger, P.L. (1994) African origin of human-specific polymorphic *Alu* insertions. *Proc Natl Acad Sci USA* **91**, 12288–12292.

Batzer, M. A., Arcot, S. S., Phinney, J. W., Algeria-Hartman, N., Kass, D. H., Milligan, S. M., Kimpton, C., Gill, P., Hochmeister, M., Ioannou, P. A., Herrera, R. J., Boudreau, D. A., Scheer, W. D., Keats, B. J. B., Deininger, P. L. & Stoneking, M. (1996) Genetic variation of recent *Alu* insertions in human populations. *J Mol Evol* **42**, 22–29.

Bowcock, A. M., Hebert, J. M., Mountain, J. L., Kidd, J. R., Rogers, J., Kidd, K. K. & Cavalli-Sforza, L. L. (1991) Study of an additional 58 DNA markers in five human populations from four continents. *Gene Geogr* **5**, 151–173.

Breeks, J. W. (1983) *Primitive tribes and monuments of the Nilgiris* Delhi: Cultural Publishing House.

Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994) *History and geography of human genes* Princeton: Princeton University Press.

Chandler, W. B. (1988) The jewel in the lotus: the Ethiopian presence in the Indus valley civilization. In: *African presence in early Asia* (eds. Sertma IV, Rashidi R.), pp. 80–105. Brunswick: NJ Transaction.

Cordaux, R., Aunger, R., Bentley, G., Nasidze, I., Sirajuddin, S. M. & Stoneking, M. (2004) Independent origins of Indian Caste and tribal paternal lineages. *Curr Biol* **14**, 231–235.

Cordaux, R., Saha, N., Bentley, G. R., Aunger, R., Sirajuddin, S. M. & Stoneking, M. (2003) Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur J Hum Genet* **11**, 253–264.

Cordaux, R. & Stoneking, M. (2003) South Asia, the Andamanese and the genetic evidence for an "early"

human dispersal out of Africa. *Am J Hum Genet* **72**, 1586–1590.

Deka, R., Jin, L., Shriver, M. D., Yu, L. M., Decroo, S., Hundrieser, J., Bunker, C. H., Ferrell, R. E. & Chakraborty, R. (1995) Population genetics of dinucleotide (dC-dA)$_n$. (dG-dT)$_n$ polymorphisms in world populations. *Am J Hum Genet* **56**, 461–474.

Endicott, P., Gilbert, M. T. P., Stringer, C., Lalueza-Fox, C., Willerslev, E., Hansen, A. J. & Cooper, A. (2003a) The genetic origins of Andaman Islanders. *Am J Hum Genet* **72**, 178–184.

Endicott, P., Macaulay, V., Kivisild, T., Stringer, C. & Cooper, A. (2003b) Reply to Cordaux and Stoneking. *Am J Hum Genet* **72**, 1590–1593.

Felsenstein, J. (1993) *PHYLIP (phylogeny inference package) version 3.5c* Seattle, WA: Department of genetics, University of Washington.

Jorde, L. B., Bamshad, M. J., Watkins, W. S., Zenger, R., Fraley, A. E., Krakowiak, P. A., Carpenter, K. D., Soodyall, H., Jenkins, T. & Rogers, A. R. (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am J Hum Genet* **57**, 523–538.

Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M. E., Ricker, C. E., Seielstad, M. T. & Batzer, M. A. (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal and Y-chromosome data. *Am J Hum Genet* **66**, 979–988.

Kidd, K. K., Morar, B., Castiglione, C. M., Zhoa, H., Pakstis, A. J., Speed, W. C., Bonne-Tamir, B., Lu, R. B., Goldman, D., Lee, C., Nam, Y. S., Grandy, D. K., Jenkins, T. & Kidd, J. R. (1998) A global survey of haplotype frequencies and linkage disequilibrium at the DRD2 locus. *Hum Genet* **103**, 211–227.

Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H.V., Stepanov, V., Gölge, M., Usanga, E., Papiha, S. S., Cinnioglu, C., King, R., Cavalli-Sforza, L., Underhill, P. A. & Villems, R. (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet* **72**, 313–332.

Majumder, P. P. (1998) People of India: biological diversity and affinities. *Evol Anthropol* **6**, 100–110.

Majumder, P. P., Roy, B., Balgir, R. S. & Dash, B. P. (1999a) Polymorphisms in the β-globin gene cluster in some ethnic populations of India and their implications on disease. In: *Molecular Intervention in Disease* New Delhi: Ranbaxy Science Foundation.

Majumder, P. P., Roy, B., Banerjee, S., Chakraborty, M., Dey, B., Mukherjee, N., Roy, M., Thakura, P. G. & Sil, S. K. (1999b) Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *Eur J Hum Genet* **7**, 435–446.

Majumder, P. & Majumder, P. P. (2000) HAPLOFREQ: A software package for maximum likelihood estimation of haplotype frequencies in populations via the EM algorithm. Tech.Rep.No.ISI/AHGU/1-2000. Indian Statistical Institute, Kolkata.

Majumder, P. P. & Mukherjee, B. N. (1993) Genetic diversity and affinities among Indian populations: An overview. In: *Human Population Genetics* (eds. P. P. Majumder), pp. 225–275. New York: Plenum.

Maloney, C. (1974) *The Races in Peoples of South Asia* New York: Holt, Rinehart & Winston.

Miller, S. A., Dykes, D. D. & Polesky, H. F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 215.

Mukherjee, N., Mitra, M., Chakraborty, M. & Majumder, P. P. (2000) Congruence of genomic and ethnolinguistic affinities among five tribal populations of Madhya Pradesh, India. *J Genet* **79**, 41–46.

Nei, M. (1972) Genetic distance between populations. *Am Nat* **106**, 283–292.

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., Antunez de Mayolo, A., Scheer, W. D., Deininger, P. L., Stoneking, M., York, D. S., Batzer, M. A. & Herrera, R.J. (1998) Polymorphic *Alu* insertions and the Asian origin of native American populations. *Hum Biol* **70**, 23–39.

Quintana-Murci, L., Semino, O., Bandelt, H. J., Passarino, G., McElreavey, K. & Santachiara-Benerecetti, A. S. (1999) Genetic evidence on an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet* **23**, 437–441.

Roychoudhury, A. K. (1982) Genetic relationships of Indian populations. In: *Human Genetics and Adaptation* (eds. K. C. Malhotra & A. Basu), pp. 147–174. New York: Plenum Press.

Roychoudhury, S., Roy, S., Basu, A., Banerjee, R., Vishwanathan, H., Usha Rani, M. V., Sil, S. K., Mitra, M. & Majumder, P. P. (2001) Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum Genet* **109**, 339–350.

Roychoudhury, S., Roy, S., Dey, B., Chakraborty, M., Roy, M., Roy, B., Ramesh, A., Prabhakaran, N., Usha Rani, M. V., Vishwanathan, H., Mitashree, M., Sil, S. K. & Majumder, P. P. (2000) Fundamental genomic unity of ethnic India revealed by analysis of mitochondrial DNA. *Curr Sci* **79**, 1182–1192.

Saha, N., Kirk, R. L., Shanbhag, S., Joshi, S. H. & Bhatia, H. M. (1974) Genetic studies among the Kadars of Kerala. *Hum Hered* **24**, 198–218.

Saha, N., Kirk, R. L., Shanbhag, S., Joshi, S. R. & Bhatia, H. M. (1976) Population genetic studies in Kerala and the Nilgiris (southwest India). *Hum Hered* **26**, 175–197.

Sherry, S. T., Harpending, H. C., Batzer, M. A. & Stoneking, M. (1997) *Alu* evolution in human populations: Using the coalescent to estimate effective population size. *Genetics* **147**, 1977–1982.

Singh, K. S. (1994) *The People of India: The scheduled tribes* Delhi: Oxford University Press.

Stoneking, M., Fontius, J. J., Clifford, S. L., Soodyall, H., Arcot, S. S., Saha, N., Jenkins, T., Tahir, M. A., Deininger, P. L. & Batzer, M. A. (1997) *Alu* insertion polymorphisms and human evolutions: evidence for a larger population size in Africa. *Genome Res* **7**, 1061–1071.

Thurston, E. (1909) *Castes and tribes of southern India,* vol IV. Madras: Government Press.

Veerraju, P., Rao, T. V., Lakshmi, N., Reshmi, S., Dey, B. & Majumder, P. P. (2001) Insertion/Deletion DNA polymorphisms in two south Indian tribal populations. *Int J Hum Genet* **1**,129–132.

Vishwanathan, H., Edwin, D., Usha Rani, M. V. & Majumder, P. P. (2003) A survey of haplotype frequencies and linkage disequilibrium at the *DRD2* locus in the Nilgiri Hills tribes, south India. *Curr Sci* **84**, 566–570.

Watkins, W. S., Ricker, C. E., Bamshad, M. J., Carroll, M. L., Nguyen, S. V., Batzer, M. A., Harpending, H. C., Rogers, A. R. & Jorde, L. B. (2001) Patterns of ancestral human diversity: an analysis of *Alu*-insertion and restriction site polymorphisms. *Am J Hum Genet* **68**, 738–752.

Wetmur, J. G., Kay, A. H., Plewinska, M. & Desnick, R. J. (1991) Molecular characterization of the human δ-aminolevulinate dehydratase 2 (ALAD^2) allele: implications for molecular screening of individuals for genetic susceptibility to lead poisoning. *Am J Hum Genet* **48**, 757–763.

Workman, P. L. & Niswander, J. D. (1970) Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. *Am J Hum Genet* **22**, 24–49.

Zischler, H., Geisert, H., von Haeseler, A. & Pääbo, S. (1995) A nuclear "fossil" of the mitochondrial D-loop and the origin of modern humans. *Nature* **378**, 489–492.