ORIGINAL ARTICLE

The Soliga, an isolated tribe from Southern India: genetic diversity and phylogenetic affinities

Diana M Morlote[1,2,5], Tenzin Gayden[1,2,5], Prathima Arvind[3], Arvind Babu[4] and Rene J Herrera[1,2]

India's role in the dispersal of modern humans can be explored by investigating its oldest inhabitants: the tribal people. The Soliga people of the Biligiri Rangana Hills, a tribal community in Southern India, could be among the country's first settlers. This forest-bound, Dravidian speaking group, lives isolated, practicing subsistence-level agriculture under primitive conditions. The aim of this study is to examine the phylogenetic relationships of the Soligas in relation to 29 worldwide, geographically targeted, reference populations. For this purpose, we employed a battery of 15 hypervariable autosomal short tandem repeat loci as markers. The Soliga tribe was found to be remarkably different from other Indian populations including other southern Dravidian-speaking tribes. In contrast, the Soliga people exhibited genetic affinity to two Australian aboriginal populations. This genetic similarity could be attributed to the 'Out of Africa' migratory wave(s) along the southern coast of India that eventually reached Australia. Alternatively, the observed genetic affinity may be explained by more recent migrations from the Indian subcontinent into Australia.
Journal of Human Genetics (2011) **56**, 258–269; doi:10.1038/jhg.2010.173; published online 10 February 2011

INTRODUCTION

India's pivotal role in the dispersal of modern humans has been supported by a number of studies.[1–10] Archeological and genetic data suggest that the country's extensive coastal area may have served as a route for human populations that migrated out of Africa ~70 000 years ago and settled Southeast Asia and Australia.[1,7,9,11–20] Anatomical similarities between some southern Indian tribes and the Australian aborigines were noticed more than a century ago by Huxley[21] who suggested an India-Australia connection. Birdsell[22] attributed the physical similarities to a possible migration of people with affinities to tribal Indians into Australia about 15 000 years ago. He hypothesized that the peopling of Australia was shaped by various migratory waves. Birdsell[22] proposed that ~15 000 years ago the Carpentarians, people who had physical characteristics similar to the Vedda Tribe of South India and Sri Lanka, arrived through the Gulf of Carpentaria and colonized northern and central Australia.

Birdsell's 'multiple migrations' hypothesis was supported by a mtDNA study[2] that argued for a recent link between aboriginal Australians and populations from the Indian subcontinent. This study was followed by Redd *et al.*[4] who reported the presence of paragroup C-M216* Y-chromosomes in both India and Australia and proposed a mid-Holocene common ancestry for these chromosomes. The genetic data suggesting multiple migrations is also supported by

changes in the Australian anthropological record between 5000 and 3000 years ago. These changes include the introduction of the dingo (Australia's wild dog), possibly arriving from India,[23] the dispersal of the Australian Small Tool tradition,[24] the appearance of technology that allowed for the processing of plants[25] and the expansion of the Pama-Nyungan language over most of Australia.[26] In addition, congruencies between the Pama-Nyungan and Dravidian languages were reported by Dixon.[27] However, a study published by Hudjashov *et al.*[28] made use of the improved resolution of the Y-chromosomal phylogeny to distinguish the Indian C sub-haplogroup (C5) from the Australian C sub-haplogroups (C4a and C4b), strongly suggesting that no migrants from India reached Australia after the original 'Out of Africa' migratory wave. In fact, the authors argue that there has been no extensive genetic contact between the first settlers of Australia and other populations, and that Australia appears to have been largely isolated since the initial migration. In addition, several studies employing autosomal markers did not show any support for a recent India-Australia connection.[12,29,30] However, a more recent study using autosomal short tandem repeats (STRs) observed significant affinity between the Arrernte people of Australia and populations from the Indian subcontinent.[31] Thus, it is evident that the available data does not provide a clear picture of when and how many times modern humans ventured into Australia via India.

[1]Department of Molecular and Human Genetics, College of Medicine, Florida International University, Miami, FL, USA; [2]Department of Biological Sciences, Florida International University, Miami, FL, USA; [3]Mary and Garry Weston Functional Genomics Unit, Thrombosis Research Institute India, Narayana Hrudayalaya Hospital, Bangalore, India and [4]RK Lakeview, Bangalore, India
[5]These authors contributed equally to the study.
Correspondence: Dr RJ Herrera, Department of Molecular and Human Genetics, College of Medicine, Florida International University, Modesto Maidique Campus, OE 304, Miami, FL 33199, USA.
E-mail: herrerar@fiu.edu

Several genetic studies[5,6,32–34] involving Indian populations have been published but they fail to reach a consensus on the origins of castes and tribes in India. A recent genome-wide study[35] employing more than 500 000 SNPs revealed that the modern Indian populations are a mixture of two source populations, the ancestral South Indians (ASI) and the ancestral North Indians (ANI). Given that tribal people represent the original inhabitants of India,[5,36–38] they are ideal candidates for genetic studies that seek to understand modern human evolution and migrations, including the peopling of Australia. The Soliga people are a tribal community found in the Biligiri Rangana (BR) Hills in the district of Chamarajanagar, in the southern state of Karnataka, India. The Soliga fit the general physical description of the Australoid ethnic group: dark complexion, curly hair, short-stature, a dolichocephalic head, a sunken nasal root and a depressed nasal bridge.[39–41] They speak *Soliganudi*[42] a dialect that has 65% lexical similarity with Kannada, a Dravidian language spoken in Karnataka, Andhra Pradesh, Tamil Nadu and Maharashtra.[43] The Soliga tribe constitutes the only scheduled tribe (Ethnic minority groups identified by the Indian Constitution for special consideration.[44]) in the BR Hills.[40] They are considered among the ancient populations of India of the 'Veddid' type (Dravidian speaking, forest dwelling, tribes of South India), believed to be true autochthones of the country.[45] Their forest-bound way of life combined with the relative inaccessibility of the BR Hills may have resulted in the cultural as well as geographic isolation of the Soligas from other populations.[40]

Autosomal STRs are hypervariable markers that have proven to be useful in the elucidation of recent human evolution.[46] Their selective neutrality, widespread distribution throughout the genome, abundance, large number of alleles and high heterozygosity provide the high resolution necessary for assessing phylogenetic relationships among closely related human populations.[46–50] In addition, unlike the uniparental mtDNA and Y-chromosome haplotypes, STRs are biparentally derived, allowing for the assessment of a more representative genetic profile of the populations under scrutiny.

In the present study, 15 autosomal STR loci were typed to characterize the genetic diversity of the Soliga people. The allelic frequencies generated were then compared with other previously published geographically targeted populations both from India and other worldwide locations. Our data indicate phylogenetic affinities between the Soligas and two Australian aboriginal populations from the Northern Territory.

MATERIALS AND METHODS

Sample collection and DNA purification
Buccal swabs were collected with informed consent from a total of 90 unrelated individuals belonging to the Soliga tribe of the BR Hills located in the Chamarajanagar district of Karnataka state in southern India (Figure 1). Regional ancestry was established by recording every person's genealogical data for at least two generations. The collection was performed in accordance with the ethical guidelines put forth by the institutions involved. DNA was extracted following manufacturer's instructions (Qiagen Inc, Valencia, CA, USA; Puregene, Gentra Systems, Minneapolis, MN, USA) and stored at $-80\,^{\circ}$C

DNA amplification and STR genotyping
DNA samples were amplified at 15 autosomal STR loci (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, THO1, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S18 and FGA). PCR amplification was performed using the AmpF*l*STR Identifiler kit[51] in an Eppendorf Mastercycler Gradient thermocycler (Eppendorf AG, Hamburg, Germany). Protocols and cycling conditions were followed as described by the manufacturer. The resulting amplicons were separated by multi-capillary electrophoresis in an ABI Prism 3100 Genetic Analyser (Applied Biosystems, Foster City, CA, USA). GeneScan 500 LIZ (Applied Biosystems) was used as an internal size standard. For genotyping,

Figure 1 Geographic locations of the Soliga tribe and previously published collections.

the GeneMapper software v3.1 (Applied Biosystems) was employed to ascertain fragment size. Alleles were designated by comparison with an allelic ladder supplied by the manufacturer.

Data analysis

Allelic frequencies were determined with the aid of the web-based GENEPOP program version 3.4.[52] The following parameters of population genetics interest were computed using the PowerStats 1.2 Software (Promega Corporation, Madison, WI, USA):[53–55] matching probability, power of discrimination, polymorphic information content, power of exclusion and typical paternity

index. Observed heterozygosity (Ho), expected heterozygosity (He), and gene diversity index were calculated with the Arlequin software package version 2.000.[56] Statistical significance was evaluated before and after applying the Bonferroni correction for 15 loci ($\alpha=0.05/15=0.0033$).

Allelic frequencies for the 30 populations listed in Table 1 were utilized to perform a correspondence analysis (CA) employing the NTSYSpc-2.02i software.[74] CA is a statistical method to analyze two-way contingency tables containing some measure of association between rows and columns. The matrix containing 30 populations arranged in columns and their corresponding allelic frequencies for 15 STR loci in rows was then fed into the software

Table 1 Populations analyzed

Population	Abbreviation	Language	Linguistic affiliations[a]	Alleles[b]	Reference
Africa					
Equatorial Guinea	EGU	Bantu	Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid	154	Alves *et al.*[57]
Angola (Cabinda)	ANG	Kongo	Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid	149	Beleza *et al.*[58]
South Africa (Cape Town)	SAF	Zulu-Bantu	Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid	144	Kido *et al.*[59]
Kenya	KEN	Swahili	Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid	139	Shephard and Herrera[49]
Mozambique (Maputo)	MOZ	Ronga	Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid	168	Alves *et al.*[57]
Tanzania	TAN	Swahili	Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid	174	Forward *et al.*[60]
Madagascar	MAD	Malagasy	Austronesian, Malayo-Polynesian, Barito, East, Malagasy	129	Regueiro *et al.*[61]
Southwest Asia					
Yemen	YEM	Arabic	Afro-Asiatic, Semitic, Central, South, Arabic	125	Shepard and Herrera[49]
Qatar	QAT	Arabic	Afro-Asiatic, Semitic, Central, South, Arabic	197	Perez-Miranda *et al.*[48]
Iraq	IRQ	Arabic	Afro-Asiatic, Semitic, Central, South, Arabic	130	Barni *et al.*[62]
Iran	IRA	Iranian	Indo-European, Indo-Iranian, Iranian	141	Shepard and Herrera[50]
South Central Asia					
Pakistan	PAK	Indo-Aryan	Indo-European, Indo-Iranian, Indo-Aryan	136	Shepard and Herrera[49]
India (Punjab)	PUN	Punjabi	Indo-European, Indo-Iranian, Indo-Aryan	134	Shepard and Herrera[49]
India (Soliga tribe)	SOL	Soliganudi (an ancient Kannada)	Dravidian, Southern	115	Present study
India (Kamma Chaudhary)	AKC	Telugu	Dravidian, South-Central, Telugu	126	Bindu *et al.*[63]
India (Kappu Naidu)	AKN	Telugu	Dravidian, South-Central, Telugu	132	Bindu *et al.*[63]
India (Santal)	SAN	Santali	Austro-Asiatic, Munda, North Munda, Kherwari, Santali	124	Banerjee *et al.*[64]
India (Oraon)	ORA	Kurukh	Dravidian, Northern, Kurukh	138	Banerjee *et al.*[64]
India (Rajbanshi)	RAJ	Bengali	Indo-European, Indo-Iranian, Indo-Aryan, Eastern zone, Bengali-Assamese	130	Roy *et al.*[65]
Bangladesh	BAN	Bengali	Indo-European, Indo-Iranian, Indo-Aryan, Eastern zone, Bengali-Assamese	142	Dobashi *et al.*[66]
Northeast Asia					
China (Shaanxi Han)	CSH	Mandarin Chinese	Sino-Tibetan, Chinese	160	Wang *et al.*[67]
Korea	KOR	Korean	Language isolate	141	Kim *et al.*[68]
Japan	JAP	Japanese	Japanese	153	Hashiyada *et al.*[69]
Southeast Asia					
Philippines	PHI	Tagalog	Austronesian, Malayo-Polynesian, Meso Philippines, central Philippine, Tagalog	128	De Ungria *et al.*[70]
Thailand	THA	Thai	Tai-Kadai, Kam-Tai, Be-Tai, Tai-Sek, Tai, Southwestern, East Central, Chiang Saeng	139	Rerkamnuaychoke *et al.*[71]
Malaysia	MAL	Malay	Austronesian, Malayo-Polynesian, Malayic, Malayian, local Malay	156	Seah *et al.*[72]
Indonesia (Java)	JAV	Indonesian	Austronesian, Malayo-Polynesian, Malayic, Malayian, local Malay	129	Shepard *et al.*[47]
Indonesia (Bali)	BAL	Indonesian	Austronesian, Malayo-Polynesian, Malayic, Malayian, local Malay	118	Shepard *et al.*[47]
Australia					
Australian Aborigines	AUS	Australian	Specific dialect not reported	161	Eckhoff *et al.*[73]
Declared Australian Aborigines	AUD	Australian	Specific dialect not reported	203	Eckhoff *et al.*[73]

[a]URL for resource http://www.ethnologue.com.
[b]Number of alleles present in the population.

for analysis. Based on the similarities and differences between the allelic frequencies, a two dimensional plot is generated displaying any association between the studied populations. The variability of the data is distributed mainly on the first and second axes, which accounts for majority of the genetic information. The method is based on a model of independent evolution and it is informative when many markers and populations are involved in a study.

Phylogenetic relationships among the populations were also assessed by using the PHYLIP v3.68 program[75] to construct a neighbor joining (NJ) dendrogram based on Nei's genetic distances. The robustness of the phylogenetic relationships was ascertained using Bootstrap analysis involving 1000 resamplings.

Inter, intra and total population genetic variance (Gst, Hs and Ht, respectively), as well as average heterozygosity for each population were computed with the DISPAN program.[76] Geographical demarcations were used to classify the 30 populations into the following eight groups: (1) sub-Sahara Africa (Madagascar, Tanzania, Mozambique, Kenya, South Africa, Angola and Equatorial Guinea); (2) Southwest Asia (Yemen, Qatar, Iraq and Iran); (3) South Central Asia (Pakistan, Rajbanshi, Oraon, Santal, Punjab, Kappu Naidu, Kamma Chaudhary, Soliga and Bangladesh); (4) South Central Asia excluding the Soliga tribe (group 4 is as group 3 minus the Soligas. The Soligas were excluded in order to assess the population's impact on genetic variance components); (5) Northeast Asia (China, Korea and Japan); (6) Southeast Asia (Thailand, Philippines, Java, Bali and Malaysia); (7) Australia (Australian aborigines and declared Australian aborigines) and (8) all populations (including all 30 populations listed in Table 1).

Genetic differences among the 30 populations were estimated by performing pairwise comparisons utilizing the Carmody program's G test.[77] A Bonferroni adjustment ($\alpha=0.05/435=0.000115$) was employed to compensate for potential type I errors.

Admixture analyses were conducted to further explore genetic affinities between the Soligas and the six groups of geographically targeted populations defined in Table 1 (as the Soliga tribe constituted the hybrid population, it was removed from the South Central Asian parental group). Although admixture tests are usually aimed at ascertaining the genetic contributions of hypothetical source populations to a hybrid, they may also reflect shared ancestry or genetic affinity in general.[61] Therefore, admixture assessment can be employed to explore relationships other than the parent-hybrid type. A second admixture test sought to investigate any possible contributions from sub-Sahara Africa, Southwest Asia, South Central Asia, Northeast Asia and Southeast Asia to the two aboriginal Australian populations was also performed. The Soligas were included in the South Central Asia parental group for this analysis.

Admixture tests were performed utilizing the weighted least squares method[48,78] with the aid of the Statistical Package for the Social Sciences (SPSS) 14.0 software (SPSS Inc., Chicago, IL, USA). The weighted least squares method was utilized as stated in the equation:

$$p_{ih} = \sum_{j=1}^{j} p_{ij} \cdot \mu_j$$

where p_{ih} is the frequency of the ith allele in the hybrid population, p_{ij} corresponds to the frequency of the ith allele in the jth reference group ($j=1,\ldots, J$), μ_j is the proportionate contribution of the jth reference gene pool to the hybrid population, and

$$\sum_{j=1}^{j} \mu_j = 1$$

RESULTS
Intra-population diversity
The allelic frequencies, expected (He) and observed (Ho) heterozygosities, Hardy–Weinberg equilibrium P-values and indices of population genetics interest (matching probability, power of discrimination, polymorphic information content, power of exclusion, typical paternity index, gene diversity index) for the Soligas are listed in Table 2. The Soliga people exhibit the lowest total number of alleles (115) when compared to the 29 reference populations in Table 1. Two

individuals presented the allelic microvariant 29.3 at locus D21S11 that has been reported elsewhere.[50,79–83] Designation was confirmed by reamplification and reanalysis. Six alleles corresponding to six different loci are present at frequencies higher than 40%. These are alleles 8 of D7S820 (0.5167), 15 of D3S1358 (0.4056), 6 of THO1 (0.4167), 19 of D2S1338 (0.4500), 13 of D19S433 (0.4556) and 17 of vWA (0.4111). Noteworthy, is the fact that the same six alleles have the highest frequencies in their corresponding loci in the Australian aborigines and declared Australian aborigines from the Northern Territory.[73] This is not the case for any of the remaining 27 reference populations.

Loci CSF1PO, D19S433 and vWA in the Soliga population were found to depart from Hardy–Weinberg equilibrium expectations at $\alpha=0.05$. However, after applying the Bonferroni correction ($\alpha=0.0033$), these deviations are rendered statistically insignificant, and no loci diverge from Hardy–Weinberg equilibrium predictions. Observed heterozygosity values were lower than He values for 8 out of the 15 loci screened. The average heterozygosity for each population is reported in Supplementary Table 1. The Soligas possess the lowest average observed heterozygosity (0.75643) of the 30 populations examined whereas the highest is observed in Madagascar (0.81237).

Intrapopulation variances (Hs) are summarized in Table 3. The sub-Saharan African populations possess the highest intrapopulation variance (0.79601) whereas the lowest is seen in the Northeast Asia group (0.77909). South Central Asia exhibits the fourth highest Hs (0.78733); however, after excluding the Soligas, intra-population diversity increases to 0.79224, the second highest overall.

Interpopulation diversity
Phylogenetic relationships among all populations were ascertained with the aid of a CA, a NJ dendogram and the Carmody program's G-test. The Gst values provided in Table 3 were generated to investigate interpopulation variance. Genetic affinities and possible contributions to hybrid populations were examined by admixture analyses.

Five main aggregates can be discerned within the CA plot (Figure 2): One consisting of all the sub-Saharan African populations except for Madagascar, a second including Northeast and Southeast Asia, a third grouping all South Central Asians except for the Soliga tribe, a fourth including the Southwest Asian populations and a fifth consisting of the two aboriginal Australian populations along with the Soligas. South Central Asia forms a rather tight cluster with the exception of Rajbanshi, which plots fairly close to the Northeast/Southeast Asia, and the Soliga tribe, which strays away from the South Central Asian cluster to join the two Australian aboriginal collections. As expected, the two southern Dravidian-speaking tribes, Kappu Naidu and Kamma Chaudhary, show more affinity to the Soligas than the other South Central Asian populations.

The NJ phylogram (Figure 3) corroborates the CA graph in that four of the clusters in the CA are clearly represented. The sub-Saharan African, Southwest Asian, Northeast and Southeast Asian populations form well delineated clades in the tree. The Soliga tribe again joins the two Australian populations, confirming their association in the CA. The South Central Asian populations segregate in between the Northeast and Southeast Asia clades found in the lower half of the dendrogram whereas the Southwest Asia and the sub-Saharan Africa clusters partition in the upper half of the phylogram.

Inter and total population variance values (Gst and Ht, respectively) are shown in Table 3. The lowest interpopulation variance value is seen in the group formed by the two aboriginal Australian populations (Gst =0.00339); however as only two Australian populations, both from Northern Territory, are used in the study, it is unlikely that the reported value is representative of the actual interpopulation variance of all

Table 2 Soliga tribe allelic frequencies ($n=90$)

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
6						0.4167									
7						0.0056									
8			0.5167			0.0833	0.1222	0.0722				0.1833			
9			0.0056			0.3056	0.2667	0.2389				0.3167	0.0167	0.0056	
9.3						0.1833									
10	0.2667		0.1111	0.2833		0.0056	0.0056	0.0833				0.0833		0.1167	
10.2										0.0167					
11	0.0944		0.2500	0.2111			0.2667	0.3500				0.3833	0.0056	0.3611	
12	0.0278		0.1056	0.3889			0.2056	0.1389		0.0500		0.0333	0.0389	0.3000	
12.2										0.0056					
13	0.0389		0.0111	0.1111			0.0889	0.1111		0.4556			0.1056	0.2056	
14	0.3444				0.0833		0.0444	0.0056		0.2333	0.0500		0.2778	0.0111	
14.2										0.0222					
15	0.1778			0.0056	0.4056					0.1056	0.0556		0.1444		
15.2										0.0722					
16	0.0056				0.2556					0.0389	0.3278		0.1500		
17	0.0333				0.1111				0.0222		0.4111		0.1833		
18	0.0111				0.1278				0.1056		0.1389		0.0056		
19					0.0167				0.4500		0.0167		0.0167		0.1167
20									0.1611				0.0056		0.1389
21									0.0333						0.1056
22									0.0778				0.0500		0.1111
23									0.1222						0.2278
24									0.0111						0.1833
25									0.0111						0.0500
26		0.0111							0.0056						0.0667
27		0.0222													
28		0.2444													
29		0.2500													
29.2		0.0556													
29.3		0.0111													
30		0.1389													
30.2		0.0056													
31		0.0278													
31.2		0.0778													
32.2		0.1167													
33.2		0.0278													
34.2		0.0111													
Ho	0.7444	0.8444	0.6222	0.6556	0.7000	0.7111	0.8111	0.7778	0.7889	0.7667	0.7222	0.6667	0.8000	0.8111	0.8889
He	0.8053	0.8379	0.6505	0.7290	0.7384	0.6963	0.8107	0.7809	0.7450	0.7209	0.7599	0.7151	0.8347	0.7589	0.8598
HWE	0.0614	0.4708	0.6140	0.0094	0.0848	0.5765	0.6975	0.5376	0.3977	0.0432	0.0491	0.9291	0.2253	0.0725	0.3332
MP	0.1094	0.0580	0.1817	0.1491	0.1067	0.1602	0.0835	0.0817	0.1002	0.1207	0.1605	0.1264	0.0578	0.1585	0.0491
PD	0.8906	0.9420	0.8183	0.8509	0.8933	0.8398	0.9165	0.9183	0.8998	0.8793	0.8395	0.8736	0.9422	0.8415	0.9509
PIC	0.7317	0.8136	0.5977	0.6598	0.6955	0.6378	0.7591	0.7456	0.7107	0.6829	0.6479	0.6605	0.8093	0.6742	0.8337
PE	0.5003	0.6839	0.3184	0.3629	0.4283	0.4457	0.6198	0.5585	0.5786	0.5387	0.4635	0.3786	0.5990	0.6198	0.7728
TPI	1.96	3.21	1.32	1.45	1.67	1.73	2.65	2.25	2.37	2.14	1.80	1.50	2.50	2.65	4.50
GDI	0.7705	0.8379	0.6505	0.7155	0.7384	0.6963	0.7951	0.7809	0.7417	0.7209	0.7023	0.7151	0.8347	0.7276	0.8559

Abbreviations: GDI, gene diversity index; He, expected heterozygosity; Ho, observed heterozygosity; HWE, Hardy–Weinberg equilibrium; MP, matching probability; PD, power of discrimination; PE, power of exclusion; PIC, polymorphic information content; TPI: typical paternity index.

aboriginal Australian populations in the continent. South Central Asia displays the highest Gst (0.01474) except for the all-populations group (Gst =0.02645). Excluding the Soliga tribe from the South Central Asia group of populations results in a 29.2% decrease in the Gst (0.01043). This considerable reduction in Gst upon the removal of the Soligas argues for the genetic uniqueness of the tribe. It is noteworthy that even with the exclusion of the Soligas, South Central Asia still has the highest Gst value of all remaining geographic groupings, indicative that

a high interpopulation diversity is inherent to the region. Total variance is highest among the sub-Saharan African group (Ht =0.80274), closely followed by the South Central Asian group without the Soliga people (Ht =0.80059), which is in turn followed by South Central Asia (Soliga people included) (Ht =0.79910).

The *G*-test results are presented in Table 4. Statistically insignificant genetic differences were observed before applying the Bonferroni correction ($\alpha=0.05$) in the following pair-wise comparisons: Kenya/

Table 3 Interpopulation and intrapopulation genetic variance

Locus	Africa			Southwest Asia			South Central Asia			South Central Asia excluding Soligas		
	Gst	Ht	Hs	Gst	Ht	Hs	Gst	Ht	Hs	Gst	Ht	Hs
D8S1179	0.00837	0.78634	0.77976	0.00590	0.83900	0.83405	0.01672	0.84152	0.82745	0.01118	0.84455	0.83511
D21S11	0.00884	0.86305	0.85543	0.00660	0.83365	0.82815	0.01064	0.85069	0.84164	0.00854	0.84994	0.84269
D7S820	0.00724	0.77193	0.76634	0.00453	0.78796	0.78439	0.02664	0.79438	0.77322	0.01616	0.80198	0.78902
CSF1PO	0.00495	0.77620	0.77235	0.00525	0.72500	0.72119	0.00827	0.70802	0.70216	0.00514	0.70460	0.70099
D3S1358	0.00759	0.73599	0.73041	0.00913	0.76190	0.75494	0.01087	0.74736	0.73923	0.00770	0.74559	0.73985
THO1	0.00460	0.73806	0.73466	0.00247	0.77523	0.77332	0.02287	0.77677	0.75901	0.01406	0.77828	0.76734
D13S317	0.01105	0.73065	0.72258	0.01368	0.77362	0.76304	0.01926	0.81114	0.79551	0.01578	0.80889	0.79612
D16S539	0.00901	0.79295	0.78581	0.00782	0.76751	0.76151	0.00887	0.79786	0.79078	0.00787	0.79884	0.79256
D2S1338	0.00692	0.88614	0.88000	0.00522	0.87889	0.87431	0.01954	0.86773	0.85078	0.00897	0.87276	0.86494
D19S433	0.01093	0.83910	0.82993	0.00900	0.83962	0.83206	0.01103	0.81210	0.80314	0.00861	0.82100	0.81393
vWA	0.00814	0.81673	0.81008	0.00509	0.80991	0.80579	0.01654	0.81186	0.79843	0.00903	0.81833	0.81094
TPOX	0.01551	0.77649	0.76444	0.00492	0.64640	0.64322	0.01563	0.71680	0.70559	0.01026	0.71220	0.70490
D18S51	0.00643	0.87534	0.86971	0.00458	0.87039	0.86641	0.00824	0.83135	0.82450	0.00672	0.82937	0.82380
D5S818	0.00980	0.77015	0.76261	0.00467	0.75803	0.75449	0.00970	0.74697	0.73972	0.01048	0.74960	0.74174
FGA	0.00674	0.88204	0.87609	0.00460	0.85954	0.85558	0.01520	0.87194	0.85869	0.01519	0.87289	0.85963
All loci	0.00838	0.80274	0.79601	0.00622	0.79511	0.79016	0.01474	0.79910	0.78733	0.01043	0.80059	0.79224

Locus	Northeast Asia			Southeast Asia			Australia			All populations		
	Gst	Ht	Hs	Gst	Ht	Hs	Gst	Ht	Hs	Gst	Ht	Hs
D8S1179	0.00396	0.84355	0.84021	0.01032	0.85539	0.84656	0.00038	0.83465	0.83433	0.02212	0.84072	0.82212
D21S11	0.00495	0.79812	0.79417	0.00849	0.84766	0.84047	0.00017	0.86818	0.86803	0.02276	0.85943	0.83987
D7S820	0.00417	0.76331	0.76013	0.00310	0.77186	0.76947	0.00054	0.75268	0.75227	0.02797	0.79193	0.76977
CSF1PO	0.00234	0.71937	0.71769	0.00562	0.71652	0.71250	0.00120	0.72645	0.72558	0.01537	0.73724	0.72591
D3S1358	0.00131	0.71486	0.71393	0.01147	0.73623	0.72779	0.00024	0.73806	0.73789	0.01402	0.74519	0.73474
THO1	0.01106	0.68538	0.67780	0.00741	0.76860	0.76290	0.00048	0.74113	0.74077	0.04637	0.78285	0.74655
D13S317	0.00256	0.80835	0.80629	0.00591	0.79197	0.78729	0.00051	0.74859	0.74821	0.04256	0.80498	0.77071
D16S539	0.00478	0.77433	0.77064	0.01164	0.78572	0.77658	0.00107	0.74543	0.74464	0.01968	0.79388	0.77826
D2S1338	0.00294	0.87244	0.86987	0.00812	0.86213	0.85513	0.00228	0.85881	0.85685	0.02415	0.88515	0.86377
D19S433	0.00669	0.79531	0.78999	0.00809	0.82747	0.82077	0.00086	0.69170	0.69110	0.02275	0.82620	0.80740
vWA	0.00169	0.79773	0.79638	0.00869	0.80874	0.80172	0.00021	0.79150	0.79133	0.02492	0.82250	0.80200
TPOX	0.00184	0.64920	0.64800	0.00967	0.60753	0.60166	0.04573	0.70749	0.67514	0.05244	0.72385	0.68590
D18S51	0.00223	0.86060	0.85867	0.00644	0.84012	0.83472	0.00028	0.87277	0.87253	0.02655	0.87211	0.84896
D5S818	0.00441	0.78999	0.78651	0.01172	0.76225	0.75332	0.00103	0.77621	0.77541	0.02580	0.77638	0.75636
FGA	0.00211	0.85786	0.85606	0.00815	0.86953	0.86244	0.00022	0.87282	0.87263	0.01249	0.87446	0.86355
All loci	0.00376	0.78203	0.77909	0.00830	0.79011	0.78355	0.00339	0.78176	0.77911	0.02645	0.80912	0.78772

Equatorial Guinea, Kenya/Angola, Qatar/Yemen and Iraq/Yemen. After applying the Bonferroni correction for potential type I errors ($\alpha=0.0001149$), several other pair-wise comparisons were also proven statistically insignificant (Table 4). The Soliga tribe was found to be statistically different from all 29 populations it was compared with. This uniqueness persisted after applying the Bonferroni correction. Statistically insignificant P-values before Bonferroni correction are shown in bold italic whereas P-values that became insignificant post Bonferroni correction are shown in bold.

Admixture proportions were calculated for the Soligas using six parental groups, which were established based on geographic divisions and phylogenetic affiliations as assessed in the CA and NJ analyses: sub-Saharan Africa, Southwest Asia, South Central Asia, Northeast Asia, Southeast Asia and Australia. Their affinities to the Soliga's gene pool are presented in Table 5. The South Central Asian group exhibits a contribution of 70.5% to the Soligas. Noteworthy, the only other group to show an affinity to the Soliga genome is the aboriginal Australian parental group, sharing 29.5% of its genetic material. When

the two Australian populations are used as hybrids (Table 6), South Central Asia is revealed as the major contributor to both the Australian aborigines and the declared Australian aborigines (58.4 and 69.1%, respectively). Interestingly, Southwest Asia and sub-Saharan Africa are the second (23.9%) and third (13.4%) major contributors, respectively, whereas Northeast Asia makes no contributions to either population, while Southeast Asia shares only 4.5% of its genetic material with the Australian aborigines and contributes nothing to the declared Australian aborigines (Table 6).

DISCUSSION

The tribal populations of India are considered the original inhabitants of the sub-continent. Therefore, it is likely that many of the unanswered questions about modern human evolution and migration can be addressed by studying the country's indigenous people. Furthermore, modern human's coastal route out of Africa that culminated in the initial settlement of Australia is thought to include migrational
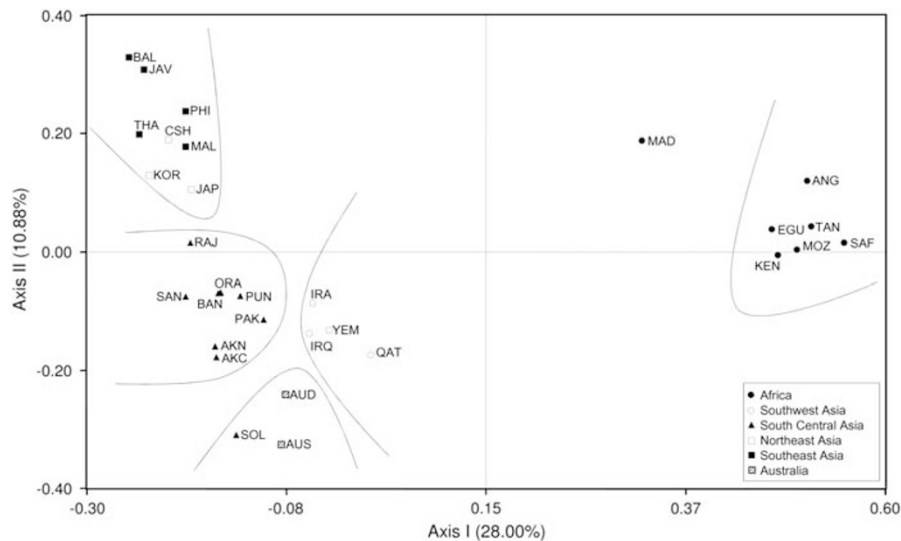
Figure 2 The correspondence analysis plot is based on 15 STR loci from the 30 studied populations. The distributions of the populations along the first two major axes were shown, which account for 28.00% (axis I) and 10.88% (axis II) of the total genetic variation.
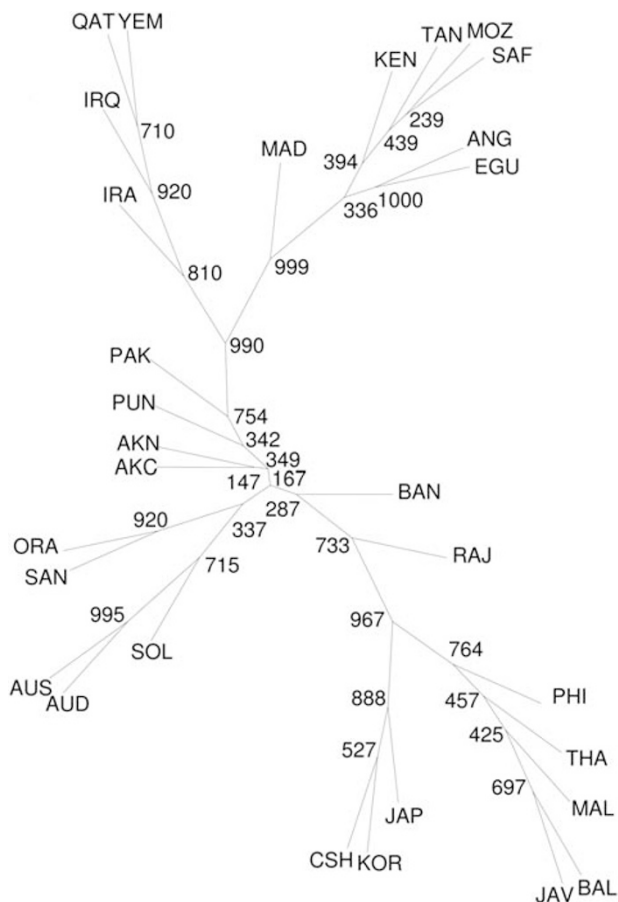

Figure 3 Neighbor joining tree based on Nei's genetic distances. The numbers at the nodes represent bootstrap values estimated from 1000 replications.

distance around the sub-continent of India, providing ample opportunity for genetic signatures to be left behind. In the present study, we assess the genetic profile of the Soligas, a southern Indian tribe, based

on 15 autosomal STR loci. In addition, we explore the tribe's phylogenetic relationships to other worldwide geographically targeted populations.

The Soligas represent a genetic isolate in the BR Hills,[40] a relatively inaccessible part of southern state of Karnataka, India. Traditionally, the Soliga do not interbreed with neighboring populations like the Kappu Naidu and Kamma Chaudhary. There is no information, to our knowledge, why the Soligas reside in this mountainous region under rather primitive conditions. Yet, in the absence of gene flow, it is possible that the Soligas have maintained a distinct gene pool. The genetic uniqueness of Soliga people is reflected in a number of population genetic parameters. For example, they possess the lowest number of alleles (115) of all the reference worldwide populations examined (Table 1). They also display the lowest average observed heterozygosity (0.75643) (Supplementary Table 1). The high degree of genetic homogeneity observed could also have been caused, in part, by their low status in the social hierarchy.

Interpopulation diversity (Gst) (Table 3) among the South Central Asian group (0.01474) is substantially higher than in the sub-Saharan African (0.00838), Southwest Asian (0.00622), Northeast Asian (0.00376), Southeast Asian (0.00830) and aboriginal Australian (0.00339) populations. The high interpopulation diversity is also reflected in the NJ tree with South Central Asian populations failing to form a distinct clade (Figure 3). The high Gst could be the result of a combination of various source populations in the peopling of South Central Asia, particularly India.[5,84] Subsequent socio-cultural barriers most likely had a role in hindering genetic flow among population groups. In order to explore the impact of the Soliga people on South Central Asia's interpopulation diversity, the tribe was excluded from the South Central Asian group, which resulted in a Gst of 0.01043, a 29.2% decrease from the original Gst, which included the Soligas. These results reflect on the genetic singularity of the Soligas. The uniqueness of the Soliga people is also evident from the G-test results wherein they were found to be statistically different from the entire set of 29 reference populations examined in this study. The Soligas exhibit significant genetic differences in relation to all 29 reference populations even after the application of the Bonferroni adjustment.

Table 4 G-test

Population		EGU	ANG	SAF	KEN	MOZ	TAN	MAD	YEM	QAT	IRQ	IRA	PAK	PUN	SOL	AKC
Equatorial guinea	EGU		224.3007	333.2830	194.7626	319.2087	334.4801	364.9931	556.7985	796.4089	776.2332	795.9909	766.8010	764.5211	1199.1587	1020.1452
Angola	ANG	**0.0182**		311.9095	202.2006	340.4551	327.7128	373.7092	565.6062	808.9791	812.0503	885.6987	824.0521	791.3032	1270.3564	1036.6266
South Africa	SAF	0.0000	0.0000		243.6518	248.7848	330.5381	385.7119	624.2535	845.4435	898.0447	916.1915	821.6796	784.9927	1184.1733	1018.1342
Kenya	KEN	**0.1709**	**0.0648**	**0.0003**		224.8122	227.2508	306.9328	448.3229	558.7856	560.1727	600.6768	594.5406	601.0364	863.3232	752.4470
Mozambique	MOZ	0.0000	0.0000	**0.0005**	**0.0039**		287.7674	370.8239	607.3222	861.5303	858.1796	917.6033	829.3272	822.9929	1189.0138	1051.3817
Tanzania	TAN	0.0000	0.0000	0.0000	**0.0378**	**0.0001**		437.2512	697.4556	1041.4496	1051.1285	1179.8137	1022.5464	972.2402	1509.6877	1347.5504
Madagascar	MAD	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		440.5806	590.6108	536.8604	524.9299	586.7352	480.4259	899.9167	685.7201
Yemen	YEM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		179.2486	153.6245	202.1445	269.4453	236.8406	626.0627	317.8607
Qatar	QAT	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0832**		243.5823	333.2791	412.2809	384.2556	909.9074	579.3433
Iraq	IRQ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.3916**	0.0000		191.0005	316.3912	309.8482	794.7230	430.0755
Iran	IRA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0091**	0.0000	**0.0337**		263.6936	255.6687	859.2436	454.9606
Pakistan	PAK	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		213.7720	586.1724	353.1689
India (Punjab)	PUN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0037**		205.7492	299.0688
India (Soligas)	SOL	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		633.7798
India (Kamma Chaudhary)	AKC	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0002**	0.0000	
India (Kappu Naidu)	AKN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
India (Santal)	SAN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
India (Oraon)	ORA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
India (Rajbanshi)	RAJ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Bangladesh	BAN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0066**	**0.0119**	0.0000	0.0000
China (Shaanxi)	CSH	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Korea	KOR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Japan	JAP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Philippines	PHI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Thailand	THA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Malaysia	MAL	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Australian aborigenes	AUS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Declared Australian aborigenes	AUD	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Java	JAV	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Bali	BAL	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 4 Continued

	AKN	SAN	ORA	RAJ	BAN	CSH	KOR	JAP	PHI	THA	MAL	AUS	AUD	JAV	BAL
EGU	943.9966	1144.7823	1122.0564	910.2997	928.9140	1449.3634	1702.5450	1999.5095	1041.8132	1488.9640	1268.4396	2152.1970	1254.5477	811.9708	1063.9211
ANG	1008.3807	1192.3932	1185.5355	918.1213	947.2019	1334.7968	1542.5803	1814.7661	987.4547	1381.0023	1227.6501	2046.9030	1261.0850	773.4238	1019.8287
SAF	954.7506	1121.2877	1173.6665	939.0475	976.4615	1417.8231	1612.9100	1860.8275	1085.5365	1482.8950	1296.7810	1861.3090	1179.0486	860.8883	1131.2350
KEN	716.1406	809.9291	811.1969	684.1375	689.8475	959.9223	1111.7213	1168.8024	782.1072	1008.0346	888.5454	1218.7571	780.6874	631.3085	836.3760
MOZ	1000.6981	1189.0138	1231.4258	956.4760	1011.5151	1571.6997	1775.4681	2070.8269	1164.1942	1666.3523	1460.6687	2285.9207	1322.6379	918.9447	836.3760
TAN	1237.2018	1482.2736	1518.7957	1141.4127	1264.8966	1953.3273	2240.3533	2849.6807	1348.9801	2042.6321	1788.5575	2891.3323	1645.7490	1022.5724	1377.5920
MAD	641.0607	757.9255	783.8444	529.9288	576.2369	775.9848	857.3165	869.1099	525.2084	731.7210	627.9091	1321.8656	816.4092	498.0347	691.6559
YEM	320.3783	468.6935	524.5886	369.3894	308.6917	525.2626	540.9344	594.0082	487.1342	456.9323	427.2851	728.1544	458.7826	401.5345	522.3990
QAT	497.4607	807.4239	860.7108	539.8386	476.9687	955.1520	995.9750	1250.2383	775.5269	924.4587	832.9911	1333.7411	754.9106	611.4866	832.6307
IRQ	403.1952	578.1790	669.3276	411.6326	379.1754	735.4686	770.7301	920.9870	598.9818	659.3774	604.6087	1103.6530	613.0706	511.5053	642.0270
IRA	396.8342	572.0316	678.6033	377.3661	327.9699	746.2988	789.9065	993.4788	552.4801	636.2730	545.6953	1372.4781	686.3104	465.4027	659.2142
PAK	283.7033	445.7553	474.6336	253.4313	216.0772	520.5606	621.9713	743.9097	524.6415	520.6755	461.1027	1136.8217	606.6910	445.0215	537.2909
PUN	236.9973	394.8783	470.1151	291.6039	205.7492	469.6430	485.3281	573.7626	399.6821	404.7916	376.7881	1024.6571	508.7344	373.1920	513.3823
SOL	586.7828	748.8760	796.1683	645.9626	608.8613	1158.7333	1142.0566	1260.6660	897.0063	1079.9656	1051.9712	1484.8560	889.9305	785.1526	905.2870
AKC	269.1855	466.2727	600.0536	346.7971	307.6448	732.8012	740.4846	945.8450	629.8422	635.4543	611.9294	1310.5170	720.0383	503.1133	658.4387
AKN	0.0000	434.8731	517.4678	301.8111	234.3697	598.0307	695.2918	835.4613	556.5879	590.0711	498.0370	1194.8398	616.4482	452.8521	601.7913
SAN	0.0000	0.0000	454.3616	407.6054	409.3986	755.5701	908.8493	1042.2404	616.0500	678.5030	616.0500	1429.8317	727.3365	446.4765	588.6795
ORA	0.0000	0.0000	0.0000	519.0915	467.8163	774.4435	919.1950	1086.4692	678.1239	454.3538	678.1239	1671.0217	907.9460	526.9359	711.6330
RAJ	0.0000	0.0000	0.0000	0.0000	227.4422	403.8986	478.0632	621.9260	426.7033	407.6015	426.7033	1179.4683	666.7315	343.9417	421.1273
BAN	**0.0008**	0.0000	0.0000	**0.0009**	0.0000	476.5106	524.9042	616.2498	463.3831	409.3986	463.3831	1225.8673	604.1436	376.2492	474.5717
CSH	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	343.3174	517.6439	380.4575	406.4190	476.1605	2415.3450	1187.3165	386.8132	508.6480
KOR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	303.5310	454.0783	461.4273	639.5113	2537.6350	1254.7749	624.2270	624.2270
JAP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	511.8138	692.6937	795.7936	1657.4690	1456.9597	467.3755	713.0462
PHI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	294.3515	286.8904	1562.2045	864.3273	241.2657	395.6574
THA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	224.9702	2149.9463	1085.2234	212.5470	379.8045
MAL	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0033**	0.0000	1972.1945	974.8283	210.9015	370.7024
AUS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	374.0680	847.0934	1139.9833
AUD	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	628.4854	833.3073
JAV	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0010**	**0.0208**	0.0000	0.0000	0.0000	234.9681
BAL	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5 Admixture analysis of the Soliga tribe using regional groups

Parental Groups	Hybrid Soliga
Northeast Asia	0.000 ± 0.113
Southeast Asia	0.000 ± 0.116
South Central Asia	0.705 ± 0.134
Southwest Asia	0.000 ± 0.100
Sub-Saharan Africa	0.000 ± 0.067
Australia	0.295 ± 0.067

Table 6 Admixture analysis of the aboriginal Australians using regional groups

	Hybrid	
Parental groups	Australian aborigines	Declared Australian aborigines
Northeast Asia	0.000 ± 0.110	0.000 ± 0.086
Southeast Asia	0.043 ± 0.112	0.000 ± 0.087
South Central Asia	0.584 ± 0.117	0.691 ± 0.091
Southwest Asia	0.239 ± 0.095	0.159 ± 0.074
Sub-Saharan Africa	0.134 ± 0.065	0.150 ± 0.050

In the CA (Figure 2), the Soliga stray from the South Central Asian cluster in the direction of the two aboriginal Australian populations. The affinity between the Soliga and the Australian groups is also supported by the NJ dendrogram (Figure 3) wherein the Soliga people form a sister clade with the two Australian collections. It is noteworthy that in the CA and the NJ tree, the Soligas exhibit stronger ties to the two aboriginal Australian populations than to the Kappu Naidu and Kamma Chaudhary, two neighboring Dravidian speaking tribes from Andhra Pradesh. Interestingly, the two Australian populations map close to the South Central Asian and Southwest Asian clusters in the CA but far from the Northeast/Southeast Asian cluster. Admixture analysis (Table 6) confirms this observation by revealing major contributions to the gene pool of the Australian aborigines and declared Australian aborigines from South Central Asia (58.4 and 69.1%, respectively) and Southwest Asia (23.9 and 15.9%, respectively). Some contribution from Africa to the aboriginal Australians is also detected (13.4% to the Australian aborigines and 15.0% to the declared Australian aborigines) but only minimal contribution is seen from Southeast Asia (4.3%) to Australian aborigines. Northeast Asia contributes nothing to the gene pool of the two aboriginal Australian populations. These results parallel an earlier study[31] that failed to detect any substantial East Asian contribution to the gene pool of the Arrernte tribe of central Australia while detecting a 56.4% Indian influence and a 25.2% Arab contribution via admixture analysis. The genetic affinity of the two aboriginal Australian populations to Southwest Asia is also reflected in the CA (Figure 2), with the two aboriginal Australian collections plotting relatively close to the Southwest Asian cluster. When the admixture test is applied using the Soliga tribe as a hybrid, the Australian group shares 29.5% of its genetic material with the Soligas, again confirming affinity between the aboriginal Australian populations and the Soliga people.

As 15 independent hypervariable autosomal STR loci were genotyped for this study, it is unlikely that the genetic affinity between the two Australian aboriginal populations and the Soligas as reflected in the admixture analyses, CA, NJ dendrogram and allelic sharing (Tables 5 and 6, and Figures 2 and 3) is the consequence of chance convergence. If indeed, as suggested by Hudjashov et al.,[28] no migrations from India reached Australia after the original settlement, then the observed affinities among the two aboriginal Australian populations and the Soligas would have to be solely attributed to the genetic signature left by the original out of Africa migrants sometime during the Pleistocene (~60 000–75 000 years ago). This early dispersal may be a genetic source for some relic populations in southern India, Southeast Asia, Papua New Guinea and Australia. These regions have been postulated to include direct descendants of the 'Out of Africa' migratory event.[30] Given that the migratory wave from India to Australia could have taken as little as ~3000 years,[7] the subsequent over 60 000 years[7] of independent evolution with no genetic exchange between Australian aborigines and the Indian populations would be expected to create substantial differences between these populations due to random genetic drift, founder effects, bottleneck events and admixture. In the case of Australian aborigines, various degrees of admixture with Europeans have occurred. It is unclear whether the sixty millennia that hypothetically separate Indian and Australian populations could have allowed for the close genetic affinity observed between the Soligas and the two Australian aboriginal populations.

Alternatively, the autosomal STR-based genetic affinity could be explained invoking additional more recent migration(s)[2,4,22] from southern India to Australia with some the Y-chromosome and mtDNA haplogroups introduced by the migrants being lost or not detected. As haploid genomes, like those contained in the Y-chromosome and mtDNA, represent one quarter of the effective population size when compared with autosomes,[85] they are more subject to genetic drift, drop out events and founder effects, which may partly account for the apparent absence of any haplogroup lineages in common between India and Australia. Even if these uniparental haplogroups were not completely lost, it is possible that any existing Indian Y lineages are so under-represented in the Indian and/or Australian gene pools that they have not been sampled or that the populations analyzed and referenced by Hudjashov et al.[28] were not impacted by the proposed recent Indian migratory event. Considering the high genetic diversity among Indian populations and the limited genetic data on Australian aboriginal populations, the above mentioned possibilities could be a likely scenarios. In relation to these conjectures, previous studies have shown a high degree of heterogeneity among the different Australian aboriginal tribes[79,86] with the most genetically distinct populations inhabiting the North Australian region.[86]

CONCLUSIONS

The present study examines the genetic profile of the Soliga tribe from Southern India based on 15 autosomal hypervariable STR loci. In addition, comparative analyses were performed to assess the phylogenetic relationship of the Soligas to a battery of worldwide geographically targeted populations. The results delineate a number of interesting genetic characteristics about the Soligas. The Soligas possess the lowest number of alleles and average observed heterozygosity when compared with all the worldwide populations examined, most likely the results of isolation and/or inbreeding. The positive effect of removing the Soligas from the South Central Asian group on the intrapopulation variance (Hs) and the negative impact on the interpopulation variance (Gst) values corroborate the genetic homogeneity of this tribe. Furthermore, the Soliga's genetic uniqueness is reflected in their statistically significant differences to all the reference populations as examined in the G-test, even after implementation of Bonferroni adjustments. Moreover, a Soliga-Australian

aboriginal genetic connection is suggested by their co-segregation in the CA and NJ analysis as well as the contribution of the Australian populations to the Soligas as detected in the admixture test. The fact that both Australian aboriginal populations from the Northern Territory and the Soligas share their six most abundant alleles (and not with any of the other examined reference populations) also suggests genetic affinities between the Soligas and the two Australian aboriginal populations. Altogether, our data portray the Soligas as a population with limited genetic diversity exhibiting unique genetic characteristics that set them apart from other populations of the Indian subcontinent. Although a recent study[28] seems to indicate that any similarities between Indian tribes and Australian aborigines are solely the result of genetic signals from the original 'Out of Africa' migration that might have taken place over 70 000 years ago, the genetic association between the Soligas and the two Northern Territory Australian aboriginal populations observed in this study suggest further inquiry into the possibility of more recent migrations from the sub-continent of India into Australia. Studies that include a larger number of tribal populations from Southern India as well as additional, better defined aboriginal Australian tribes would likely shed more light on a possible recent India-Australia connection.

ACKNOWLEDGEMENTS

1 Quintana-Murci, L., Semino, O., Bandelt, H. J., Passarino, G., McElreavey, K. & Santachiara-Benerecetti, A. S. Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat. Genet.* **23,** 437–441 (1999).
2 Redd, A. J. & Stoneking, M. Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua new Guinean populations. *Am. J. Hum. Genet.* **65,** 808–828 (1999).
3 Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408,** 708–713 (2000).
4 Redd, A. J., Roberts-Thomson, J., Karafet, T., Bamshad, M., Jorde, L. B., Naidu, J. M. et al. Gene flow from the Indian subcontinent to Australia evidence from the Y chromosome. *Curr. Biol.* **12,** 673–677 (2002).
5 Basu, A., Mukherjee, N., Roy, S., Sengupta, S., Banerjee, S., Chakraborty, M. et al. Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res.* **13,** 2277–2290 (2003).
6 Cordaux, R., Aunger, R., Bentley, G., Nasidze, I., Sirajuddin, S. M. & Stoneking, M. Independent origins of Indian caste and tribal paternal lineages. *Curr. Biol.* **14,** 231–235 (2004).
7 Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W. et al. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308,** 1034–1036 (2005).
8 Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A. G., Singh, V. K., Rasalkar, A. A. et al. Reconstructing the origin of Andaman islanders. *Science* **308,** 996 (2005).
9 Sun, C., Kong, Q. P., Palanichamy, M., Agrawal, S., Bandelt, H. J., Yao, Y. G. et al. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol. Biol. Evol.* **23,** 683–690 (2006).
10 Majumder, P. P. The human genetic history of South Asia. *Curr. Biol.* **20,** 184–187 (2010).
11 Kennedy, K. A. R., Deraniyagala, S. U., Roertgen, W. J., Chiment, J. & Disotell, T. Upper Pleistocene fossil hominids from Sri Lanka. *Am. J. Phys. Anthropol.* **72,** 441–461 (1987).
12 Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton University Press, Princeton, NJ, USA, 1994).
13 Lahr, M. M. & Foley, R. Multiple dispersals and modern human origins. *Evol. Anthropol.* **3,** 48–60 (1994).
14 Stringer, C. Palaeoanthropology: coasting out of Africa. *Nature* **405,** 24–27 (2000).
15 Cann, R. L. Genetic clues to dispersal in human populations: retracing the past from the present. *Science.* **291,** 1742–1748 (2001).
16 Maca-Meyer, N., González, A. M., Larruga, J. M., Flores, C. & Cabrera, V. M. Major genomic mitochondrial lineages delineate early human expansions. *BMC. Genet.* **2,** 1471–2156 (2001).
17 Lewin, R. & Foley, R. *Principles of Human Evolution* 2nd edn.(Blackwell Publishing, MA, USA, 2004).
18 Palanichamy, M., Sun, C., Agrawal, S., Bandelt, H. J., Kong, Q. P. & Khan, F. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of south Asia. *Am. J. Hum. Genet.* **75,** 966–978 (2004).
19 Tanaka, M., Cabrera, V. M., Gonzalez, A. M., Larruga, J. M., Takeyasu, T., Fuku, N. et al. Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome. Res.* **14,** 1832–1850 (2004).
20 Kong, Q. P., Bandelt, H. J., Sun, C., Yao, Y. G., Salas, A. & Achilli, A. Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum. Mol. Genet.* **15,** 2076–2086 (2006).
21 Huxley, T. H. On the geographical distribution of the chief modifications of mankind. *J. Ethnol. Soc. London.* **2,** 404–412 (1870).
22 Birdsell, J. B. *Microevolutionary Patterns in Aboriginal Australia: A Gradient Analysis of Clines* (Oxford University Press, New York, 1993).
23 Gollan, K. Prehistoric dogs in Australia: an Indian origin? in *Recent Advances in Indo-Pacific Prehistory* (eds Misra, V.N. & Bellwood, P.) 439–443 (Oxford and IBH Publishing Co, New Delhi, 1985).
24 Glover, I. C. & Presland, G. Microliths in Indonesian flaked stone industries in *Recent Advances in Indo-Pacific Prehistory* (eds Misra, V.N. & Bellwood, P.) 185–195 (Oxford and IBH Publishing Co, New Delhi, 1985).
25 Beaton, J. M. Dangerous harvest: investigations in the late prehistoric occupation of upland south-east central Queensland PhD Thesis, Australian National University, Canberra (1977).
26 Evans, N. & Jones, R. The cradle of the Pama-Nyungans: archaeological and linguistic speculations in *Archaeology and Linguistics: Aboriginal Australia in Global Perspective* (eds McConvell, P & Evans, N.) 385–417 (Oxford University Press, Melbourne, 1997).
27 Dixon, R. M. W. *The Languages of Australia* (Cambridge University Press, New York, 1980).
28 Hudjashov, G., Kivisild, T., Underhill, P. A., Endicott, P., Sanchez, J. J., Lin, A. A. et al. Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc. Natl Acad. Sci.* **104,** 8726–8730 (2007).
29 Kirk, R. L. & Thorne, A. G. *The Origin of the Australians* (Humanities Press, New Jersey, 1976).
30 Nei, M. & Roychoudhury, A. K. Evolutionary relationships of human populations on a global scale. *Mol. Biol. Evol.* **10,** 927–943 (1993).
31 Alfonso-Sanchez, M. A., Perez-Miranda, A. M. & Herrera, R. J. Autosomal microsatellite variability of the Arrernte people of Australia. *Am. J. Hum. Biol.* **20,** 91–99 (2008).
32 Bamshad, M., Kivisild, T., Watkins, W. S., Dixon, M. E., Ricker, C. E., Rao, B. B. et al. Genetic evidence on the origins of Indian caste populations. *Genome. Res.* **11,** 994–1004 (2001).
33 Sahoo, S., Singh, A., Himabindu, G., Banerjee, J., Sitalaximi, T., Gaikwad, S. et al. A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc. Natl Acad. Sci.* **103,** 843–848 (2006).
34 Sengupta, S., Zhivotovsky, L. A., King, R., Mehdi, S. Q., Edmonds, C. A., Chow, C. E. et al. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am. J. Hum. Genet.* **2,** 202–221 (2006).
35 Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature.* **461,** 489–494 (2009).
36 Thapar, R. *A History of India: Volume 1* (Penguin Books, London, 1966).
37 Ray, N. *Nationalism in India* (Aligarh Muslim University, Aligarh, India, 1973).
38 Majumder, P. P. Genomic inferences on peopling of south Asia. *Curr. Opin. Genet. Dev.* **18,** 280–284 (2008).
39 Chopra, P. N. *The Gazetteer of India* (Ministry of Education and Social Welfare, India, 1965).
40 Morab, S. G. *The Soliga of Biligiri Rangana Hills* (Anthropological Survey of India, India, 1977).
41 Majumder, P. P. People of India: biological diversity and affinities. *Evol. Anthropol.* **6,** 100–110 (1998).
42 Zaraska, N. A. *Health Behaviors of the Soliga Tribe Women* Master's Thesis, Queen's University, Canada 1997.
43 Gordon, R. G. *Ethnologue: Languages of the World* 15th edn. (SIL International, Texas, 2005).
44 Sujatha, K. Education among scheduled tribes in *India Education Report* 1st edn (eds Govinda, R.) 87–94 (Oxford University Press, New Delhi, 2002).
45 Sarkar, S. S. *The Aboriginal Races of India* 1st edn. (Bookland Ltd, Calcutta, 1954).
46 Rowold, D. J. & Herrera, R. J. Inferring recent human phylogenies using forensic STR technology. *Forensic Sci. Int.* **133,** 260–265 (2003).
47 Shepard, E. M., Chow, R. A., Suafo'a, E., Addison, D., Perez-Miranda, A. M., Garcia-Bertrand, R. L. et al. Autosomal STR variation in five Austronesian populations. *Hum. Biol.* **77,** 825–851 (2005).
48 Perez-Miranda, A. M., Alfonso-Sanchez, M. A., Pena, J. A. & Herrera, R. J. Qatari DNA variation at a crossroad of human migrations. *Hum. Hered.* **61,** 67–79 (2006).
49 Shepard, E. M. & Herrera, R. J. Genetic encapsulation among Near Eastern populations. *J. Hum. Genet.* **51,** 467–476 (2006a).
50 Shepard, E. M. & Herrera, R. J. Iranian STR variation at the fringes of biogeographical demarcation. *Forensic Sci. Int.* **158,** 140–148 (2006b).
51 Applied Biosystems. *AmpFlSTR Identifiler PCR Amplification Kit User's Manual* (Applied Biosystems, Foster City, CA, USA, 2001).
52 Raymond, M. & Rousset, F. Genepop (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* **86,** 248–249 (1995).
53 Jones, D. A. Blood samples: probability of discrimination. *J. Forensic Sci. Soc.* **12,** 355–359 (1972).
54 Brenner, C. H. & Morris, J. W. *Proceedings for the International Symposium on Human Identification* (Promega Corporation, Madison, 1990).
55 Tereba, A. *Profiles in DNA* (Promega Corporation, Madison, 1999).

56 Schneider, S., Kueffer, J. M., Roessli, D. & Excoffier, L. *Arlequin v. 2.000: A Software for Population Genetic Analysis* (Genetics and Biometry Laboratory, University of Geneva, Geneva, 2000).

57 Alves, C., Gusmão, L., Damasceno, A., Soares, B. & Amorim, A. Contribution for an African autosomic STR database (AmpF/STR Identifiler and Powerplex 16 system) and a report on genotypic variations. *Forensic Sci. Int.* **139,** 201–205 (2004).

58 Beleza, S., Alves, C., Reis, F., Amorim, A., Carracedo, A. & Gusmão, L. 17 STR data (AmpF/STR Identifiler and Powerplex 16 system) from Cabinda (Angola). *Forensic Sci. Int.* **141,** 193–196 (2004).

59 Kido, A., Dobashi, Y., Fujitani, N., Hara, M., Susukida, R., Kimura, H. *et al.* Population data on the AmpFISTR identifiler loci in Africans and Europeans from South Africa. *Forensic Sci. Int.* **168,** 232–235 (2007).

60 Forward, B. W., Eastman, M. W., Nyambo, T. B. & Ballard, R. E. AMPF/STR® identifiler™ STR allele frequencies in Tanzania, Africa. *J. Forensic Sci.* **53,** 245–247 (2008).

61 Regueiro, M., Mirabal, S., Lacau, H., Caeiro, J. L., Garcia-Bertrand, R. L. & Herrera, R. J. Austronesian genetic signature in east African Madagascar and Polynesia. *J. Hum. Genet.* **53,** 106–120 (2008).

62 Barni, F., Berti, A., Pianese, A., Boccellino, A., Miller, M. P., Caperna, A. *et al.* Allele frequencies of 15 autosomal STR loci in the Iraq population with comparisons to other populations from the middle-eastern region. *Forensic Sci. Int.* **167,** 87–92 (2007).

63 Bindu, H. G., Trivedi, R. & Kashyap, V. K. Allele frequency distribution based on 17 STR markers in three major Dravidian linguistic populations of Andhra Pradesh, India. *Forensic Sci. Int.* **170,** 76–85 (2007).

64 Banerjee, J., Trivedi, R. & Kashyap, V. K. Polymorphism at 15 short tandem repeat AmpFISTR Identifiler TM loci in three aboriginal populations of India: an assessment in human identification. *J. Forensic Sci.* **50,** 1229–1234 (2005).

65 Roy, S., Eaaswarkhanth, M., Dubey, B. & Haque, I. Autosomal STR variations in three endogamous populations of West Bengal, India. *Leg. Med.* **10,** 326–332 (2008).

66 Dobashi, Y., Kido, A., Fujitani, N., Hara, M., Susukida, R. & Oya, M. STR data for the AmpFLSTR Identifiler loci in Bangladeshi and Indonesian populations. *Leg. Med.* **7,** 222–226 (2005).

67 Wang, Z. Y., Yu, R. J., Wang, F., Li, X. S. & Jin, T. B. Genetic polymorphisms of 15 STR loci in Han population from Shaanxi (NW china). *Forensic Sci. Int.* **147,** 89–91 (2005).

68 Kim, Y. L., Hwang, J. Y., Kim, Y. J., Lee, S., Chung, N. G., Goh, H. G. *et al.* Allele frequencies of 15 STR loci using AmpF/STR Identifiler kit in a Korean population. *Forensic Sci. Int.* **136,** 92–95 (2003).

69 Hashiyada, M., Itakura, Y., Nagashima, T., Nata, M. & Funayama, M. Polymorphism of STRs by multiplex analysis in japanese population. *Forensic Sci. Int.* **133,** 250–253 (2003).

70 De Ungria, M. C. A., Roby, R. K., Tabbada, K. A., Rao-Coticone, S., Tan, M. M. M. & Hernandez, K. N. Allele frequencies of 19 STR loci in a Philippine population generated using AmpFISTR multiplex and ALF singleplex systems. *Forensic Sci. Int.* **152,** 281 (2005).

71 Rerkamnuaychoke, B., Rinthachai, T., Shotivaranon, J., Jomsawat, U., Siriboonpiputtana, T., Chaiatchanarat, K. *et al.* Thai population data on 15 tetrameric STR loci— D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818 and FGA. *Forensic Sci. Int.* **158,** 234–237 (2006).

72 Seah, L. H., Jeevan, N. H., Othman, M. I., Jaya, P., Ooi, Y. S., Wong, P. C. *et al.* STR data for the AmpF/STR Identifiler loci in three ethnic groups (Malay, Chinese, Indian) of the Malaysian population. *Forensic Sci. Int.* **138,** 134–137 (2003).

73 Eckhoff, C., Walsh, S. J. & Buckleton, J. S. Population data from sub-populations of the northern territory of Australia for 15 autosomal short tandem repeat (STR) loci. *Forensic Sci. Int.* **171,** 237–249 (2007).

74 Rohlf, F. *NTSYSpc* (Exter Publishing, Setauket, NY, 2002).

75 Felsentein, J. *Phylogeny Inference Package (PHYLIP) Version 3.6a3* (Department of Genetics, University of Washington, Seattle, 2002).

76 Ota, T. *DISPAN: Genetic Distance and Phylogenetic Analysis* (Institute of Molecular Evolutionary Genetics, Pennsylvania State University, Pennsylvania, 1993).

77 Carmody, G. *G-test* (Carleton University, Ottawa, 1990).

78 Long, J. C., Williams, R. C., McAuley, J. E., Medis, R., Partel, R., Tregellas, W. M. *et al.* Genetic variation in Arizona Mexican Americans: estimation and interpretation of admixture proportions. *Am. J. Phys. Anthropol.* **84,** 141–157 (1991).

79 Ayres, K. L., Chaseling, J. & Balding, D. J. Implications for DNA identification arising from an analysis of Australian forensic databases. *Forensic Sci. Int.* **129,** 90–98 (2002).

80 Allor, C., Einum, D. D. & Scarpetta, M. Identification and characterization of variant alleles at CODIS STR loci. *J. Forensic Sci.* **50,** 1128–1133 (2005).

81 Ferdous, A., Eunus Ali, M., Alam, S., Hossai, T., Hany, U., Dissing, J. *et al.* Genetic data on 10 autosomal STR loci in the Bangladeshi population. *Leg. Med.* **8,** 297–299 (2006).

82 Ali, M. E., Ferdous, A., Alam, S., Hany, U., Hossain, T., Hasan, M. *et al.* Identification of variant alleles at AmpFISTR SGM plus STR loci in a sample population of Bangladesh. *Afr. J. Biotechnol.* **7,** 3603–3605 (2008).

83 Short Tandem Repeat DNA Internet Database, http://www.cstl.nist.gov/biotech/strbase/.

84 Kashyap, V. K., Guha, S., Sitalaximi, T., Bindu, G. H., Hasnain, S. E. & Trivedi, R. Genetic structure of Indian populations based on fifteen autosomal microsatellite loci. *BMC. Genet.* **7,** 28 (2006).

85 Jobling, M. A. & Tyler-Smith, C. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* **4,** 598–612 (2003).

86 Walsh, S. J., Mitchell, R. J., Watson, N. & Buckleton, J. S. A comprehensive analysis of microsatellite diversity in Aboriginal Australians. *J. Hum. Genet.* **52,** 712–728 (2007).