Legal Medicine xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Legal Medicine

journal homepage: www.elsevier.com/locate/legalmed

Announcement of Population Data

Genetic variation of 15 autosomal microsatellite loci in a Tamil population from Tamil Nadu, Southern India

Kuppareddi Balamurugan [a,*], S. Kanthimathi [b], M. Vijaya [b], G. Suhasini [b], George Duncan [c], Martin Tracey [d], Bruce Budowle [e,f]

[a] School of Criminal Justice, University of Southern Mississippi, 118 College Drive, # 5127, Hattiesburg, MS 39406, USA
[b] Department of Genetics, Postgraduate Institute of Basic Medical Sciences, University of Madras, Chennai 600 113, India
[c] DNA Laboratory, Broward County Sheriff's Office, 201 SE 6th Street, Fort Lauderdale, FL 33301, USA
[d] Department of Biological Sciences, Florida International University, Miami, FL 33199, USA
[e] Department of Forensic and Investigative Genetics, University of North Texas Health Science Center, Ft Worth, TX 76107, USA
[f] Institute of Investigative Genetics, University of North Texas Health Science Center, Ft Worth, TX 76107, USA

ARTICLE INFO

Article history:
Received 23 February 2010
Received in revised form 18 May 2010
Accepted 27 July 2010
Available online xxxx

Keywords:
STR
Tamil population
South Indian population
DNA frequency estimate
Gene diversity
Genetic distance

ABSTRACT

The genetic profiles for 15 autosomal microsatellite loci were analyzed in a Tamil population from Southern India to study the genetic diversities and relatedness of this population with other national and global populations. Statistical analyses of the data revealed all loci were within Hardy–Weinberg Equilibrium (HWE) expectations with the exception of the locus D5S818 ($p = 0.011$). A significantly greater inter-individual variation (Fst = 99%) observed within the individuals among the four subgroups in this study and low population differentiation (Fst = 1%) suggests relative genetic closeness of these four subgroups. This indicates that the populations in the southern region of India might have a common ancestry or probably experienced high gene flow during the period of their coexistence. The Neighbor Joining tree derived from genetic distances of samples from this study and other national and global populations show clustering of all the Indian populations in one branch of the tree while the African and Middle Eastern populations cluster in a separate branch. Principal Co-ordinate Analysis of the genetic distance data show clustering similar to the NJ tree.

Populations: Samples (136 individuals) from a Tamil population from the state of Tamil Nadu, India were analyzed for 15 autosomal STR markers. The population samples consist of four subgroups and all of them speak the same Dravidian language (Tamil) and reside in the state of Tamil Nadu. Because of the homogeneity of genetic data in 3 out of 4 of the subgroups, all the samples were pooled and treated as one group. The map of Tamil Nadu and some of the anthropological information of ten populations of Tamil Nadu are described elsewhere [1].

DNA extraction: Blood samples were collected with informed consent from 136 unrelated individuals. Samples were collected in accordance with the ethical guidelines stipulated by the institutions involved in this study. DNA was isolated following the method as described in Miller et al. [2]. Total human DNA was quantitated using the Quantifiler® human DNA quantitation kit as per manufacturer's recommendations (Applied Biosystems, Foster City, CA).

PCR amplification: Approximately 1–2 ng of DNA template was amplified using the AmpF*l*STR® Identifiler™ PCR amplification kit (Applied Biosystems, Foster City, CA) at the following STR loci: D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, CSF1PO, FGA, THO1, TPOX and vWA.

Typing: Amplicons were separated by capillary electrophoresis in an ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA) and sized with Genescan-500 LIZ size standard. Allele calls and genotyping were carried out using GeneMapper® ID v3.2 software (Applied Biosystems, Foster City, CA).

Quality control: Positive and negative controls as specified in the Identifiler™ kit user's manual. The data were analyzed and verified by two independent analysts.

Analyses of data: The STR allelic frequencies and other forensic parameters, including power of discrimination (PD), matching probability (MP), and power of exclusion (PE) were calculated using PowerStats program version 1.2 [3]. Heterozygosities and HWE *p*-values were calculated using Arlequin V3.1 [4]. Nei's genetic distances [5] were derived from allele frequencies of the samples from this study as well as other regional and global populations listed in Table 1 using POPTREE2 [6]. These genetic distances were utilized to generate a Neighbor Joining (NJ) dendrogram using the same program [6]. The robustness of the phylogenetic relationships established by the NJ tree was assessed using bootstrap analysis with 1000 replications. Graphical representation of genetic distances (Dst) of the population in this study along with 16 other

* Corresponding author. Tel.: +1 601 266 6048; fax: +1 601 266 4391.
E-mail address: Kuppareddi.Balamurugan@usm.edu (K. Balamurugan).

1344-6223/$ - see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.legalmed.2010.07.004

Table 1
Population data used for analysis using NJ tree and PCA plot from genetic distances.

	Population	Abbreviation	Number of loci	References
1	Bahrain, Middle East	BAH	15	Shepard and Herrera [15]
2	Chakkiliyar, South India	CHA	13	Sitalaximi et al. [16]
3	Egypt, Africa	EGY	15	Shepard and Herrera [15]
4	Georgia, Asia	GEO	15	Shepard and Herrera [15]
5	Gounder, South India	GOU	13	Sitalaximi et al. [16]
6	Irular, South India	IRU	13	Sitalaximi et al. [16]
7	Jordan, Middle East	JOR	15	Shepard and Herrera [15]
8	Kenya, Africa	KEN	15	Shepard and Herrera [15]
9	Oman, Middle East	OMN	15	Shepard and Herrera [15]
10	Pakistan, Asia	PAK	15	Shepard and Herrera [15]
11	Pallar, South India	PAL	13	Sitalaximi et al. [17]
12	Punjabi, North India	PUN	15	Shepard and Herrera [15]
13	Rwandan(Hutu), Africa	RWH	15	Shepard and Herrera [15]
14	Sudan, Africa	SUD	15	Shepard and Herrera [15]
15	Tamil, South India	MC	15	Balamurugan et al. [18]
16	Tamil, South India	TAM	15	Present study
17	Yemen, Middle East	YEM	15	Shepard and Herrera [15]

Table 2
Allele frequencies of 15 autosomal STR loci for the Tamil population (n = 272 chromosomes).

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
6						0.324									
7			0.040			0.118	0.015							0.004	
8	0.004		0.217			0.107	0.250	0.114				0.232			
9	0.018		0.059	0.007		0.305	0.107	0.125				0.221		0.026	
9.1			0.004												
9.3						0.140									
10	0.173		0.246	0.243		0.007	0.059	0.132				0.096	0.015	0.136	
11	0.066		0.210	0.313			0.224	0.272		0.004		0.415	0.033	0.342	
12	0.088		0.202	0.353			0.243	0.195		0.066		0.029	0.022	0.283	
12.2										0.015					
13	0.096		0.018	0.070			0.063	0.147		0.335		0.007	0.158	0.195	
13.2										0.011					
14	0.184		0.004	0.007	0.055		0.037	0.011		0.217	0.177		0.272	0.015	
14.2										0.059					
15	0.195			0.004	0.283		0.004	0.004		0.118	0.114		0.191		
15.2										0.092					
16	0.143			0.004	0.290				0.007	0.055	0.206		0.110		
16.2										0.026					
17	0.033				0.228				0.066	0.004	0.257		0.074		
18					0.136				0.221		0.132		0.055		
19					0.007				0.136		0.099		0.048		0.085
19.2															0.004
20									0.092		0.015		0.011		0.140
21									0.0368						0.188
21.1															0.004
22									0.070				0.004		0.129
22.2															0.007
23									0.232				0.004		0.180
23.2															0.004
24									0.063				0.004		0.140
25									0.067						0.059
26									0.007						0.048
27															0.015
28		0.136													
29		0.191													
30		0.154													
30.2		0.022													
31		0.055													
31.2		0.114													
32		0.011													
32.2		0.184													
33.2		0.110													
34.2		0.022													
Ho	0.904	0.831	0.802	0.721	0.662	0.765	0.816	0.838	0.860	0.809	0.743	0.794	0.838	0.699	0.860
He	0.858	0.861	0.805	0.717	0.765	0.760	0.811	0.823	0.854	0.810	0.823	0.718	0.843	0.748	0.867
HWE	0.731	0.917	0.364	0.336	0.135	0.791	0.912	0.136	0.285	0.590	0.640	0.184	0.994	0.011*	0.710
MP	0.044	0.038	0.074	0.142	0.092	0.099	0.068	0.064	0.046	0.066	0.056	0.142	0.044	0.115	0.038
PD	0.956	0.962	0.926	0.858	0.908	0.901	0.932	0.936	0.954	0.934	0.944	0.858	0.956	0.885	0.962
PIC	0.840	0.840	0.770	0.660	0.720	0.720	0.780	0.800	0.830	0.780	0.790	0.670	0.820	0.700	0.850
PE	0.804	0.658	0.602	0.461	0.372	0.535	0.629	0.672	0.715	0.616	0.497	0.588	0.672	0.426	0.715

Ho: observed heterozygosity; He: expected heterozygosity; HWE: Hardy–Weinberg Equilibrium p-values; MP: matching probability; PD: power of discrimination; PIC: polymorphic information content; PE: power of exclusion.
* Deviation from HWE.

Please cite this article in press as: Balamurugan K et al. Genetic variation of 15 autosomal microsatellite loci in a Tamil population from Tamil Nadu, Southern India. Leg Med (2010), doi:10.1016/j.legalmed.2010.07.004

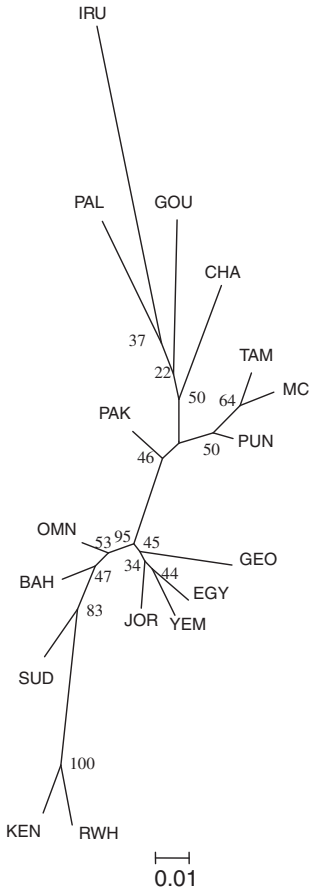

Fig. 1. Neighbor Joining (NJ) tree based on Nei's genetic distances generated from allele frequencies of the autosomal microsatellite loci. The numbers at the nodes represent bootstrap values estimated from 1000 replications. The population codes and descriptions are given in Table 1.

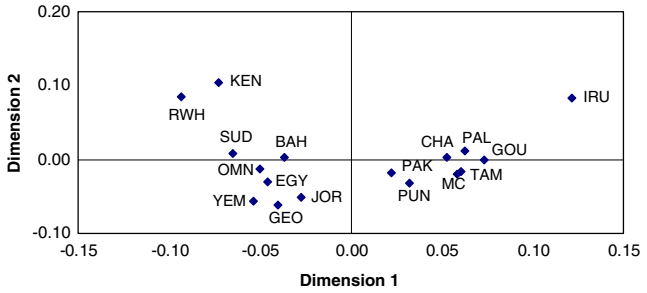

Fig. 2. Principal Co-ordinate Analysis (PCA) of genetic distances derived from the population from this study and other regional, national and global populations. The population codes and descriptions are given in Table 1.

national and global populations was performed based on Principal Co-ordinate Analysis (PCA) plot using GenAlEx software package version 6.3 [7]. Analysis of Molecular Variance (AMOVA) was also calculated using the same software [7]. The homogeneity among the subgroups was assessed with the G-test program [8].

Results: The population used in this study as well as other populations used for measuring genetic distances and comparisons are listed in Table 1. Allele frequency data of the 15 STR loci are presented in Table 2. These STR loci were found to be highly polymorphic and statistical analysis of these data revealed all loci met Hardy–Weinberg Equilibrium (HWE) expectations with the exception of D5S818 ($p = 0.011$). The combined power of discrimination and the probability of exclusion for all the loci are >0.99999 while the combined matching probability is 3×10^{17} and the average heterozygosity is 0.7961. Analysis of Molecular Variance (AMOVA) of the four subgroups included in this study showed 99% variance within individuals and 1% variance among populations signifying the homogeneity among the four subgroups. These results have also been supported from other study [1] involving 10 different caste groups from Tamil Nadu, India. Recently, Watkins et al. [9] studied several caste populations spanning different social hierarchy from Tamil Nadu using autosomal STRs, Y chromosomal SNPs, and mitochondrial haplotypes and concluded that the Tamil south Indian castes are only modestly differentiated from one another with 0.96% of STR variance. Estimates for heterozygosity and repeat variance in these populations indicated no substantial between-caste differences or excess homozygosity in these caste groups. The genotypes of all the 136 samples studied are provided as Supplementary data (Table 3).

Other remarks: Global and regional comparisons, based on allele frequencies with diverse populations from Southern and Northern India, Pakistan, Middle East and Africa were performed to address relationships with neighboring and distant populations following a hypothesis of an out of Africa theory following a route through Southern India [10,11]. The genetic relationships of these populations were demonstrated in a Neighbor Joining dendrogram (Fig. 1) using Nei's genetic distance [5]. Examination of the dendrogram shows clustering of all the Indian populations as one group and other Middle Eastern and African populations as another group with the Pakistani population in between but closer to the Indian populations. Closer affinity can be seen among the Middle Eastern populations while the three African populations branched away from the Middle Eastern group, whereas the Indian populations show more diversity among themselves. Principal Co-ordinate Analysis of the genetic distance data show clustering of most of the Indian populations (except Irular, South India) while other African and Middle Eastern populations segregated as another group (Fig. 2). The grouping of populations in the PCA plot is consistent with the clustering pattern of the NJ tree. India and its populations have received wide attention from population geneticists, anthropologists and archaeologists due to their rich history, biological diversity, caste system, consanguineous marriages and genetic isolation due to the caste hierarchical system [12,13]. In spite of the linguistic homogeneity in Southern India, the cultural barriers created by the caste system that may prevent some gene flow between populations could be a partial contributor for genetic diversification among the people of the southern region [14].

Conflict of interest: The authors declare no financial or other conflict of interest.

Acknowledgements

The authors wish to thank Dr. A. Ramesh, Chairman, Department of Genetics, University of Madras, India for sharing the samples for this study. This study was partly supported by a startup grant for one of the authors (KB) from the University of Southern Mississippi, Hattiesburg, MS. Partial financial support was provided to the Department of Genetics, University of Madras by the Department of Biotechnology, Government of India. We thank Ms. Nicole Mullins, School of Criminal Justice, University of Southern Mississippi for technical assistance and data analysis.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.legalmed.2010.07.004.

References

[1] Kanthimathi S, Vijaya M, Ramesh A. Genetic study of Dravidian castes of Tamil Nadu. Supplementary electronic data. J Genet 2008;87:175–9.
[2] Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.
[3] Tereba A. Tools for analysis of population statistics. Profiles DNA 1999;2:14–6.
[4] Excoffier L, Laval G, Schneider S. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 2005;1:47–50.
[5] Nei M. Genetic distance between populations. Am Nat 1972;106:283–91.
[6] Takezaki N, Nei M, Tamura K. POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with windows-interface. Mol Biol Evol 2009. doi:10.1093/molbev/msp31.
[7] Peakall R, Smouse PE. GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes 2006;6:288–95.
[8] Carmody G. G-test software package: department of biology. Ottawa, CA: Carleton University; 1991.
[9] Watkins WS, Thara R, Mowry BJ, Zhang Y, Witherspoon DJ, Tolpinrud W, et al. Genetic variation in South Indian castes: evidence from Y-chromosome, mitochondrial, and autosomal polymorphisms. BMC Genet 2008;9:86.
[10] Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, et al. Reconstructing the origin of Andaman Islanders. Science 2005;308:996.
[11] 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 2005;308:1034–6.
[12] Reich D, Thangaraj K, Patterson N, Price A, Singh L. Reconstructing Indian population history. Nature 2009;461:489–94.
[13] Petraglia M, Korisettar R, Boivin N, Clarkson C, Ditchfield P, Jones S, et al. Middle Paleolithic assemblages from the Indian subcontinent before and after the Toba super-eruption. Science 2007;317:114–6.
[14] Majumder PP. People of India: biological diversity and affinities. Evol Anthropol 1998;6:100–10.
[15] Shepard EM, Herrera RJ. Genetic encapsulation among near eastern populations. J Hum Genet 2006;51:467–76.
[16] Sitalaximi T, Trivedi R, Kashyap VK. Autosomal microsatellite profile of three socially diverse ethnic Tamil populations of India. J Forensic Sci 2003;48:211–4.
[17] Sitalaximi T, Trivedi R, Kashyap VK. Genotype profile for thirteen tetranucleotide repeat loci and two pentanucleotide repeat loci in four endogamous Tamil population groups of India. J Forensic Sci 2002;47:1168–73.
[18] Balamurugan K, Prabakaran N, Duncan G, Budowle B, Tahir M, Tracey M. Allele frequencies of 13 STR loci and the D1S80 locus in a Tamil population from Madras, India. J Forensic Sci 2001;46:1515–7.

Please cite this article in press as: Balamurugan K et al. Genetic variation of 15 autosomal microsatellite loci in a Tamil population from Tamil Nadu, Southern India. Leg Med (2010), doi:10.1016/j.legalmed.2010.07.004