Genetic Polymorphisms and Forensic Efficiency of 16 X-chromosomal STR Loci for Sri Lankan Population

Nandika Perera
Genetech (Sri Lanka)
Gayani Galhena (gayanihg@gmail.com)
Department of Zoology and Environment Sciences, University of Colombo, Colombo 03, Sri Lanka
Gaya Ranawaka
Open University of Sri Lanka

Research Article

Keywords: short tandem repeat (STR), multiplex PCR system, Indo-Aryan ancestry

Posted Date: February 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-149443/v1

Abstract

A new 16 X- short tandem repeat (STR) multiplex PCR system has recently been developed for Sr Lankans, though its applicability in evolutionary genetics and forensic investigations has not been thoroughly assessed. In this study, 838 unrelated individuals covering all four major ethnic groups (Sinhalese, Sri Lankan Tamils, Indian Tamils and Moors) in Sri Lanka were successfully genotyped using this new multiplex system. The results indicated a high forensic efficiency for the tested loci in all four ethnicities confirming its suitability for forensic applications of Sri Lankans. Allele frequency distribution of Indian Tamils showed subtle but statistically significant differences from those of Sinhalese and Moors, in contrast to frequency distributions previously reported for autosomal STR alleles. This suggest a sex biased demographic history among Sri Lankans requiring a separate X-STR allele frequency database for Indian Tamils. Substantial differences observed in the patterns of LD among the four groups demand the use of a separate haplotype frequency databases for each individual ethnicity. When analysed together with other 14 world populations, all Sri Lankan ethnicities except Indian Tamils clustered closely with populations from Indian Bhil tribe, Bangladesh and Europe reflecting their shared Indo-Aryan ancestry.

Introduction

Sri Lanka is an island country in South Asia, located in the Indian Ocean, close to India. Due to its strategic position at middle of the maritime silk route from China to Europe, it was well known to the outside world from ancient times as a trading hub. The diverse ethnicities that compose the 20 million population inhabiting the island as per the last population census[1] have descended mainly from numerous groups of migrants who came to the island at various historical time periods. Their overpowering impact have confined the original inhabitants of the country to a few dry zone areas, forming a tribal group known as the Veddahs (aboriginals)[2], represented by about 10,000 individuals[3].

Today, the Sinhalese make the largest ethno-cultural group in Sri Lanka, having a population of 15.17 million (74.9% of the total population)[1]. The Sinhalese make a unique population in the world as the only ethnic group that speaks Sinhala, a branch of the Indo-European (Indo-Aryan) language family[4]. According to historical chronicles, the Bengali prince Vijaya and his seven hundred followers, who are descendants from the Indo-Aryans natives of the northern Indian subcontinent laid the foundation to the Sinhalese in 543 BCE[4]. They vanquished the aboriginal Veddas and converted Sri Lanka to a Sinhalese territory until the Dravidian rulers from South India invaded the northern part of the island in the fifth century AD[4]. Since then, there had been frequent migrationsF by South Indians into the country, which gave rise to the second largest ethnic group in Sri Lanka known as the "Ceylon Tamil" (2.27 million people representing 11.2%)[1].

A third ethnic group was established in the country when Arab traders visiting the country for commercial purposes settled in Sri Lanka in 1000 AD, leading to intermarriages with the Sinhalese and the Sri Lankan Tamils[4]. The group now known as Moors comprise 9.2%[1] (1.86 million) of the Sri Lankan population and maintain unique sociocultural features which are based largely on the Islamic faith. They speak a Dravidian language that contains large number of Arabic words that is generally referred to as “Arabic Tamil”[5]. However, some scholars attribute the origin of Moors to South Indian traders, who later settled in Sri Lanka[5]. This view in part is based on the similarities shared by Sri Lankan Moors with the Tamil Muslims of Tamil Nadu. Indian Tamils make the 4th largest ethnic group in Sri Lanka and comprise the descendants from plantation workers brought to Sri Lanka from South India by the English rulers who colonized Sri Lanka in 19th century[4]. Comprising a relative minority of 0.84 million people (4.2% of the total population)[1], Indian Tamils are chiefly confined to the central hills in Sri Lanka with a relatively low admixture with other ethnic groups due to socio cultural reasons associated with their more recent immigrant status.

In addition to these ethnicities, around 0.5% of the Sri Lankan population comprises other minor ethnic groups belonging to numerous descents. They include Malays (descendants from island of Java) Burghers (descendants of colonists from Portugal, Netherlands and UK) and other Chinese and African migrants who came to the island in the 18th and 19th centuries.

Because of the demographic history and geneflow, the genetic position occupied by each of these ethnic groups, both at local and global scales is not clear. The wide array of genetic markers currently available provide an opportunity for reliably elucidating their genetic affinities. However, depending on the type of genetic markers used, the ancestral information of populations deduced from the analyses could be quite different. On the other hand, each of these different approaches can complement each other with their characteristic genetic information. Although a few previous studies that had been conducted to understand the genetic substructure and underlying heterogeneity of Sri Lankan ethnicities using autosomal[6], Y chromosomal[7] and mitochondrial markers[8], none has utilized those present on the X chromosome. X-chromosome markers, particularly short tandem repeats (STRs), with the advantageous features of both autosomal and uniparental biomarkers, play an important role in evolutionary studies[9,10] as well as in forensic genetics[11].They assist in the interpretation of complex kinship cases on its own or in conjunction with other marker like autosomal STRs. Analysis of X chromosome STRs (X-STRs) is specially advantageous in complex cases, where at least one female is involved, such as in a deficient paternity case of a female child, cases involving female siblings sharing a common biological father, questioned relationships between paternal grandmother-granddaughter or other distant female relatives[12]. They are also of special importance in forensic case work in which female traces are to be identified in male background contamination[12,13]. The use of clusters of tightly linked X-STRs forming highly informative haplotypes is particularly effective in such cases[13]. In addition, X-chromosome markers have a proven utility in tracing the sex biased demographic history among populations with complex admixture and geneflow patterns. Consequently, X-chromosome markers have gained significant importance in population and forensic genetic studies in the past two decades. However, the routine practice of molecular genetics in Sri Lanka presently comprise only of autosomal, Y-chromosomal and mitochondrial DNA analyses. The scope of X-chromosome markers is yet to be investigated for Sri Lankans.

Recognizing this vital need exist in the field of molecular genetics in Sri Lanka, we recently developed a multiplex X-STR system with 16 X-STR markers distributed from 9.198 Mb to 149.460 Mb of the X chromosome[14] with the aim of incorporating X-STR analysis in to molecular forensic and evolutionary genetics investigations in the country. Thirteen of these STR markers are in four closely linked clusters (each spanning < 3 cM) that are likely to produce stable haplotypes (Cluster I; DXS10148-DXS10135-DXS8378 (Xp22), Cluster II; DXS7132-DXS10079-DXS10074-DXS10075 (Xq12), Cluster III: DXS6801-DXS6809-DXS6789 (Xq21), Cluster IV; DXS7424-DXS101-DXS7133 (Xq22)). Additional three unlinked markers were also included from both p (DXS9902 at Xp22) and q (HPRTB at Xq26 and DXS7423 at Xq28) arms to have a better coverage of the X chromosome. The assay was validated for the Sinhalese using 200 unrelated individuals of which 120 were males. In the present study, we extended the analysis to evaluate the forensic efficiency of this novel 16 X-STR assay to all four major ethnicities in Sri Lanka (Sinhalese, Sri Lankan Tamil, Indian Tamils and Moors) and constructed an allele and haplotype frequency database for Sri Lankans for forensic and kinship analysis purposes. Further, we have conducted a comprehensive analysis of the possible linkage disequilibrium (LD) among the selected X-STR markers, which is essential in making valid conclusions on existing relationships. Here, we report for the first time, the X-STR based population genetic information of all four main ethnicities in Sri Lanka. In addition, Pairwise genetic distances based on Fst were also calculated between the Sri Lankan population and populations from South, South East and East Asia, Europe, Africa and Brazil based on the data extracted from literature to elucidate the genetic substructure between Sri Lankan and other global populations.

Results And Discussion

Polymorphism

We typed 16 X-STR loci for 838 unrelated individuals of the Sri Lankan population covering the four major ethnicities. Complete DNA profiles were obtained for all male samples without any allele dropouts. Altogether 203 alleles were observed for all four populations with 30 private alleles (Sinhalese: total alleles (A_T) = 184; private alleles (A_p) = 17, Sri Lankan Tamils: A_T=164; A_p=5, Indian Tamils: A_T=156; A_p=5 and Moors: A_T=150; A_p=3). When female samples were tested for the conformity to Hardy-Weinberg equilibrium, no significant deviations were observed for any of the tested loci after adjusting for multiple comparisons (Bonferroni corrected P = 0.0031) (Supplementary Table S1). The exact test of population differentiation did not detect significant differences in allele distribution among the male and female samples and hence the allele frequencies were combined for both sexes for further analysis. The allele frequencies of the four ethnicities are displayed in Supplementary Tables S2-S5.

The forensic parameters calculated based on the combined allele frequency data (Tables 1–8) indicated high values in general for all 16 markers. For example, expected heterozygosity (He) values were above 0.6 for all X-STR markers in all four ethnicities, except for DXS7423 in the two Tamil ethnicities. PIC values of 14 of the studied markers were above 0.6 for all ethnicities, among which, eight showed values above 0.7. Even the two markers, which showed PIC < 0.6, had it approximated to 0.5 (DXS7423: 0.4989 for Indian Tamils) or higher (DXS7133: >0.5 for all ethnicities). The power of discrimination for males (PD_m) ranged from 0.5840 to 0.9380, while for females (PD_f) ranged from 0.7385 to 0.9926. However, MEC values showed a broader variation among the markers; i.e. MEC in deficient cases (MEC_{Kru}) ranged from 0.2958 to 0.8739, MEC in normal trios ($MEC_{Des-trio}$) from 0.4956 to 0.9345 and MEC in duos ($MEC_{Des-duo}$) from 0.3548 to 0.8803. Among all the tested loci, DXS10135 showed the highest value for all forensic parameters, suggesting it to be the most informative marker for Sri Lankans, while DXS7423 showed the lowest and the least informative. It is also interesting to note that the least polymorphic marker, DXS7423, had two of its alleles (alleles 14 and 15) presented in about 80% of the studied samples. Nevertheless, the combined power of discrimination for the 16 tested loci in both males (CPD_m >0.999999999958569) and females (CPD_f=1.000000000000000), as well as combined MEC indices calculated for deficiency ($CMEC_{Kru}$>0.999998849524199), normal trio ($CMEC_{Des-trio}$>0.999999999678351), and duo cases ($CMEC_{Des-duo}$ >0.999999800357383) were equally high for all four ethnicities (Table 9a).

Table 1
Forensic parameters of 16 X-STR loci among the four ethnic populations.

	DXS10148				DXS10135			
	Sinhala	**SLT**	**INT**	**Moors**	**Sinhala**	**SLT**	**INT**	**Moors**
PIC	0.8889	0.8838	0.8831	0.8802	0.9260	0.9345	0.9210	0.9206
He	0.8976	0.8929	0.8923	0.8898	0.9304	0.9380	0.9258	0.9255
PD_f	0.9808	0.9794	0.9792	0.9782	0.9907	0.9926	0.9896	0.9895
PD_m	0.8976	0.8929	0.8923	0.8898	0.9304	0.9380	0.9258	0.9255
MEC_{Kru}	0.7940	0.7865	0.7850	0.7803	0.8586	0.8739	0.8498	0.8491
$MEC_{Des-trio}$	0.8889	0.8838	0.8831	0.8802	0.9260	0.9345	0.9210	0.9206
$MEC_{Des-duo}$	0.8082	0.8010	0.7997	0.7956	0.8663	0.8803	0.8583	0.8577

Table 2
Forensic parameters of 16 X-STR loci among the four ethnic populations.

	DXS8378				DXS9902			
	Sinhala	**SLT**	**INT**	**Moors**	**Sinhala**	**SLT**	**INT**	**Moors**
PIC	0.6526	0.6285	0.6044	0.6314	0.6927	0.7085	0.6633	0.6837
He	0.7020	0.6835	0.6496	0.6874	0.7384	0.7511	0.7099	0.7326
PD_f	0.8618	0.8449	0.8320	0.8463	0.8859	0.8954	0.8692	0.8796
PD_m	0.7020	0.6835	0.6496	0.6874	0.7384	0.7511	0.7099	0.7326
MEC_{Kru}	0.4555	0.4266	0.4119	0.4287	0.4991	0.5187	0.4699	0.4863
$MEC_{Des-trio}$	0.6526	0.6285	0.6044	0.6314	0.6927	0.7085	0.6633	0.6837
$MEC_{Des-duo}$	0.5082	0.4825	0.4563	0.4857	0.5523	0.5702	0.5201	0.5423

Table 3
Forensic parameters of 16 X-STR loci among the four ethnic populations.

	DXS7132				DXS10079			
	Sinhala	SLT	INT	Moors	Sinhala	SLT	INT	Moors
PIC	0.6831	0.6873	0.7026	0.7002	0.79741	0.7873	0.8040	0.7962
He	0.7283	0.7311	0.7456	0.7427	0.82046	0.8125	0.8233	0.8207
PD_f	0.8810	0.8838	0.8923	0.8913	0.94471	0.9396	0.9495	0.9433
PD_m	0.7283	0.7311	0.7456	0.7427	0.82046	0.8125	0.8233	0.8207
MEC_{Kru}	0.4912	0.4966	0.5134	0.5111	0.64806	0.6329	0.5350	0.6433
$MEC_{Des-trio}$	0.6831	0.6873	0.7026	0.7002	0.79741	0.7873	0.8040	0.7962
$MEC_{Des-duo}$	0.5419	0.5465	0.5639	0.5610	0.67993	0.6670	0.6910	0.6777

Table 4
Forensic parameters of 16 X-STR loci among the four ethnic populations.

	DXS10074				DXS10075			
	Sinhala	SLT	INT	Moors	Sinhala	SLT	INT	Moors
PIC	0.7746	0.7881	0.7830	0.7956	0.6341	0.6230	0.6699	0.6888
He	0.8021	0.8136	0.8088	0.8181	0.6846	0.6735	0.7123	0.7301
PD_f	0.9333	0.9397	0.9376	0.9444	0.8500	0.8429	0.8748	0.8858
PD_m	0.8021	0.8136	0.8088	0.8181	0.6846	0.6735	0.7123	0.7301
MEC_{Kru}	0.6145	0.6325	0.6277	0.6474	0.4372	0.4260	0.4824	0.5040
$MEC_{Des-trio}$	0.7746	0.7881	0.7830	0.7956	0.6341	0.6230	0.6699	0.6888
$MEC_{Des-duo}$	0.6509	0.6677	0.6619	0.6780	0.4885	0.4766	0.5278	0.5493

Table 5
Forensic parameters of 16 X-STR loci among the four ethnic populations.

	DXS6801				DXS6809			
	Sinhala	SLT	INT	Moors	Sinhala	SLT	INT	Moors
PIC	0.6219	0.6311	0.6390	0.6342	0.7769	0.7876	0.7923	0.7655
He	0.6759	0.6768	0.6943	0.6892	0.8006	0.8107	0.8141	0.7933
PD_f	0.8409	0.8498	0.8512	0.8484	0.9365	0.9410	0.9437	0.9294
PD_m	0.6759	0.6768	0.6943	0.6892	0.8006	0.8107	0.8141	0.7933
MEC_{Kru}	0.4201	0.4380	0.4374	0.4309	0.6234	0.6369	0.6436	0.6038
$MEC_{Des-trio}$	0.6219	0.6311	0.6390	0.6342	0.7769	0.7876	0.7923	0.7655
$MEC_{Des-duo}$	0.4750	0.4848	0.4940	0.4882	0.6544	0.6678	0.6735	0.6397

Table 6
Forensic parameters of 16 X-STR loci among the four ethnic populations.

	DXS6789				DXS7424			
	Sinhala	SLT	INT	Moors	Sinhala	SLT	INT	Moors
PIC	0.7644	0.7577	0.7562	0.7773	0.8063	0.7766	0.7430	0.8072
He	0.7917	0.7876	0.7853	0.8029	0.8288	0.8047	0.7776	0.8298
PD_f	0.9293	0.9250	0.9248	0.9355	0.9482	0.9337	0.9159	0.9484
PD_m	0.7917	0.7876	0.7853	0.8029	0.8288	0.8047	0.7776	0.8298
MEC_{Kru}	0.6035	0.5921	0.5937	0.6215	0.6587	0.6143	0.5684	0.6588
$MEC_{Des-trio}$	0.7644	0.7577	0.7562	0.7773	0.8063	0.7766	0.7430	0.8072
$MEC_{Des-duo}$	0.6385	0.6300	0.6292	0.6549	0.6909	0.6526	0.6118	0.6916

Table 7
Forensic parameters of 16 X-STR loci among the four ethnic populations.

	DXS101				DXS7133			
	Sinhala	SLT	INT	Moors	Sinhala	SLT	INT	Moors
PIC	0.7955	0.8251	0.8167	0.7976	0.6028	0.5630	0.5600	0.5968
He	0.8178	0.8437	0.8364	0.8204	0.6523	0.6028	0.6027	0.6558
PD_f	0.9445	0.9569	0.9534	0.9448	0.8296	0.8025	0.7995	0.8225
PD_m	0.8178	0.8437	0.8364	0.8204	0.6523	0.6028	0.6027	0.6558
MEC_{Kru}	0.6478	0.6895	0.6769	0.6488	0.4089	0.3773	0.3721	0.3939
$MEC_{Des-trio}$	0.7955	0.8251	0.8167	0.7976	0.6028	0.5630	0.5600	0.5968
$MEC_{Des-duo}$	0.6780	0.7166	0.7053	0.6803	0.4556	0.4139	0.4110	0.4496

Table 8
Forensic parameters of 16 X-STR loci among the four ethnic populations. PIC: polymorphism information content, He: expected heterozygosity, PD female: power of discrimination in females, PD male: power of discrimination in males, MEC Krü: mean exclusion chance Kruger, MEC Des.trio: mean exclusion chance Desmaris trio. MEC Des.duo: mean exclusion chance Desmaris duo, SLT: Sri Lankan Tamil, INT: Indian Tamil.

	HPRTB				DXS7423			
	Sinhala	SLT	INT	Moors	Sinhala	SLT	INT	Moors
PIC	0.7078	0.6790	0.6713	0.6912	0.5466	0.5146	0.4956	0.5788
He	0.7487	0.7263	0.7179	0.7358	0.6220	0.5983	0.5840	0.6460
PD_f	0.8959	0.8778	0.8738	0.8856	0.7817	0.7549	0.7385	0.8074
PD_m	0.7487	0.7263	0.7179	0.7358	0.6220	0.5983	0.5840	0.6460
MEC_{Kru}	0.5212	0.4849	0.4774	0.5011	0.3432	0.3127	0.2958	0.3775
$MEC_{Des-trio}$	0.7078	0.6790	0.6713	0.6912	0.5466	0.5146	0.4956	0.5788
$MEC_{Des-duo}$	0.5698	0.5375	0.5288	0.5514	0.4020	0.3720	0.3548	0.4336

Table 9
Combined forensic efficiency parameters calculated for the 16 X-STR loci in four ethnic group based only on allele frequency data (a) and on both allele frequency and haplotype frequency data for those loci in LD (b)

a) Based on individual X-STR				
	Sinhalese	SLT	INT	Moors
CPD_f	1.000 000 000 000 000	1.000 000 000 000 000	1.000 000 000 000 000	1.000 000 000 000 000
CPD_m	0.999 999 999 976 606	0.999 999 999 985 151	0.999 999 999 958 569	0.999 999 999 980 686
$CMEC_{Kru}$	0.999 999 404 082 778	0.999 999 642 120 508	0.999 998 849 524 199	0.999 999 430 187 790
$CMEC_{Des-trio}$	0.999 999 999 808 127	0.999 999 999 886 900	0.999 999 999 678 351	0.999 999 999 831 577
$CMEC_{Des-duo}$	0.999 999 865 132 380	0.999 999 915 282 522	0.999 999 800 357 383	0.999 999 876 826 033
b) Based on X-STR + relevant linkage haplotype				
	Sinhalese	SLT	INT	Moors
CPD_f	1.000 000 000 000 000	1.000 000 000 000 000	1.000 000 000 000 000	1.000 000 000 000 000
CPD_m	0.999 999 999 957 448	0.999 999 999 985 151	0.999 999 999 939 787	0.999 999 999 973 317
$CMEC_{Kru}$	0.999 999 830 029 405	0.999 999 642 120 508	0.999 999 040 493 880	0.999 999 555 022 837
$CMEC_{Des-trio}$	0.999 999 999 809 273	0.999 999 999 886 900	0.999 999 999 616 777	0.999 999 999 811 582
$CMEC_{Des-duo}$	0.999 999 938 607 489	0.999 999 915 282 522	0.999 999 810 368 648	0.999 9 889 885 194

CPD_f: combined power of discrimination in females, CPD_m: combined power of discrimination in males, CMECKru: combined mean exclusion chance Kruger, $CMEC_{Des-trio}$: combined mean exclusion chance Desmaris trio, $CMEC_{Des-duo}$: combined mean exclusion chance Desmaris duo, SLT: Sri Lankan Tamil, INT: Indian Tamil

For those loci which were found to be in LD, it is more appropriate to calculate the forensic efficiency based on the observed haplotype frequencies instead of individual allele frequencies, since each haplotype is assumed to behave as an allele. Accordingly, forensic efficiency was recalculated for the X-STR loci for those ethnicities which exhibited LD (see section below on LD and haplotypes) based on their particular haplotype frequencies (Table 10). The combined forensic efficiency parameters for the 16 X-STR were also recalculated based on the efficiency of both individual markers and haplogroups (for those markers in LD). For this purpose, the 16 markers were treated as 12 loci for Sinhalese, while they were treated as 15 loci for Indian Tamils and Moors. For Sri Lankan Tamils, recalculation was not necessary as there was no LD detected within the 16 loci tested. As shown in Table 9b, the combined forensic

efficiency parameters calculated based on the presence of LD, yielded values of same order of magnitude compared to what was yielded based on individual allele frequency data; i.e. CPD_f=1.000000000000000; CPD_m> 0.999999999939787; $CMEC_{Kru}$> 0.999999040493880; $CMEC_{Des-trio}$ > 0.999999999616777; $CMEC_{Des-duo}$ > 0.999999810368648. These results suggest that the 16 tested X-STR loci are appropriate candidates for kinship and forensic analysis among the four Sri Lankan ethnicities, especially with the cases involving female offspring.

Table 10

Forensic statistical parameters of the four haplogroups. PIC: polymorphism information content, He: expected heterozygosity, PD_f: power of discrimination in females, PD_m: power of discrimination in males, MEC_{Kru}: mean exclusion chance Kruger, $MEC_{Des-trio}$: mean exclusion chance Desmaris trio. $MEC_{Des-duo}$: mean exclusion chance Desmaris duo, INT: Indian Tamil.

Haplotype	Ethnic group	PIC	He	HD	PD_f	PD_m	MEC_{Kru}	$MEC_{Des-trio}$	$MEC_{Des-duo}$
DXS10135-DXS8378	Sinhala	0.9730	0.9736	0.9774	0.9987	0.9736	0.9478	0.9730	0.9484
DXS7132-DXS10074-DXS10075	Sinhala	0.9753	0.9758	0.9796	0.9989	0.9758	0.9536	0.9753	0.9526
DXS7424-DXS101	Sinhala	0.9608	0.9621	0.9659	0.9973	0.9621	0.9250	0.9608	0.9264
DXS10074-DXS10075	INT	0.9147	0.9201	0.9350	0.9882	0.9201	0.8393	0.9147	0.8484
	Moors	0.9247	0.9289	0.9430	0.9907	0.9289	0.8576	0.9247	0.8649

Population Differentiation Among Sri Lankan Ethnicities

A locus-by-locus analysis of molecular variance (AMOVA) was carried out including all 16 loci, grouping the four ethnicities based on their linguistic origins (Table 11) to understand the extent of genetic differentiation among them. Out of the four populations, Sinhalese are known to have an Indo-Aryan origin, which is different from the Dravidian linguistic origin of the other three ethnicities. However, a significant variation was not detected among the two linguistic groups (Fct= -0.00059; P > 0.05) though a subtle, but statistically significant variation was detected among populations within groups (Fsc = 0.0018; P = 0.0108). The global AMOVA results as a weighted average over loci showed that most of the variance in the samples is attributable to within-individual variation (97.97%) and between ethnic group variation is around 0.18%. To better understand this observed population structure, pairwise comparisons (pairwise Fst analysis) were carried out among all four ethnicities (Table 12). According to the Fst values obtained, Indian Tamils were shown to have a subtle but statistically significant genetic subdivision from Sinhalese (Fst = 0.0029; P = 0.0000) and Moors (Fst = 0.0038; P = 0.0000) while Sinhalese, Sri Lankan Tamils and Moors are highly panmictic (P > 0.05). Further, the two Tamil ethnicities were shown to share a common genetic background (P > 0.05).

Table 11
Analysis of molecular variance (AMOVA) among the four ethnicities based on linguistic groups

Source of variation	Sum of squares	Variance components	Percentage of variation	Fixation Indices	P value
Among groups	8.456	-0.00363 Va	-0.06	FCT : -0.00059	0.80156
Among populations within groups	16.965	0.01091 Vb	0.18	FSC : 0.00178	0.01075
Among individuals within populations	3758.554	0.11718 Vc	1.91	FIS : 0.01913	0.00098
Within individuals	3641.500	6.00908 Vd	97.97	FIT : 0.02029	0.00000
Total	7425.474	6.13354			

Table 12
Pairwise Fst and P-values values for four ethnicities Below diagonal: Pairwise Fst values, Above diagonal: P-values, Significant P-Values are indicated in bold font

	Sinhala	SL Tamil	IND Tamil	Moors
Sinhala		0.18919 ± 0.0370	**0.00000 ± 0.0000**	0.36036 ± 0.0450
SL Tamil	0.00062		0.17117 ± 0.0394	0.08108 ± 0.0252
IND Tamil	0.00294	0.00082		**0.00000 ± 0.0000**
Moors	0.00033	0.00134	0.00375	

Since phylogenetic trees constructed from genetic distances can easily deduce the evolutionary relationships and origins of different populations[15,16], UPGMA method was applied for the four ethnicities to further evaluate their genetic affinities. As shown in the phylogram (Fig. 1), Sinhalese and Moors are genetically closely associated with each other and also with Sri Lankan Tamils to a lesser extent. Although Indian Tamil group is placed at a distant position from Moors and Sinhalese, the close genetic affinity between the two Tamil groups are apparent in the phylogram. Nei genetic distances for the six pairwise ethnic groups are listed in the Table 13.

Table 13
Nei genetic distances among different pairwise ethnic groups.

Pairwise population	Nei genetic distances
Sinhala-SL Tamil	0.0174
Sinhala-IN Tamil	0.0205
Sinhala-Moors	0.0123
SL Tamil-IN Tamil	0.0237
SL Tamil-Moors	0.0233
IN Tamil-Moors	0.0293

These findings agree with the historical data on early settlement of the four ethnic groups in Sri Lanka. According to anthropological and archaeological evidence, Sri Lankan Tamils have a very long history in Sri Lanka and have lived in the island since at least around the second century BCE. They have arrived in Sri Lanka from various parts of the Indian subcontinent, either with their original families or alone, and subsequently uniting with the Sinhalese through matrimonial bonds. Indian Tamils on the other hand were brought to Sri Lanka to work in estates during the British colonization and had minimum admixture with the native Sinhalese or with Sri Lankan Tamils, who were more economically independent at the time and had better social status. Even at present, majority of Indian Tamils are congregated around the plantation estates of central hill area of the country forming a separate community. In contrast, Sri Lankan Moors have descended exclusively from Muslim male merchants of either Arabic or of Indian origin[5], who came to Sri Lanka for trading. During the 14th century, they started to settle in coastal areas in Sri Lanka and espoused local women, who were either Sinhalese or Sri Lankan Tamil. Thus, it is not surprising to see this local female ancestry reflected among Moors via our X-STR analysis, despite their Arabic origin, in the light that the X chromosome spends two third of its lifetime within females. Alternatively, the genetic similarity observed between Sri Lankan Tamils and Moors may be reflecting the Indian origin of Moors as some scholars claims it to be[5]. Likewise, the genetic similarity observed between the two Tamil populations might also lie in their common Indian origin.

Since the X chromosome reflect more of the maternal blood line, results reported in mitochondrial (mt) DNA studies are also of much relevance to the present context. Although there are no mtDNA studies conducted on Moors, research on the other three ethnic groups have reported the existence of a very fine genetic structure with a more closer genetic relationship among Sinhalese and Sri Lankan Tamils, in comparison to that among Sinhalese and Indian Tamils[8]. When taken together with the fact that autosomal STR analysis have failed in detecting a genetic structure among the four ethnic groups[6], our results suggest a sex-biased demographic history for Sri Lankan ethnicities.

Genetic distance between Sri Lankan ethnicities and other world populations

In order to understand the genetic composition of Sri Lankan ethnicities in relation to other global populations, allele frequencies generated in the present study were compared with 17 other world populations. Depending on the availability of data in the literature, 8–16 common X-STR markers were used for the analysis. Accordingly, three populations from South Asia (Bhil tribe and Brahmin caste in India[17,18], Bangladesh[19] and Pakistan[20,21]), two from South East Asia (Malaysians[22]and Thailand[23]), three from East Asia (China[24–26], Japan[24,27,28] and Taiwan[29,30]), five from Europe (Germany[31–34], Italy[35–39], Sweden[40], Denmark[41] and North Portugal[42]) two from Africa (Somalia[41], and Ivory Coast[43]) and Brazil[44] were compared (Supplementary Table S6).

Among the three South Asian populations, Bhil tribe and Brahmin caste populations of India are the most geographically proximal populations to Sri Lankan ethnics groups. Both these Indian populations did not show a significant genetic differentiation with the Sri Lankan ethnicities` with respect to any of the tested loci. Similarly, Bangladesh population also did not exhibit any population subdivision with Sri Lankan ethnicities. However, Pakistan population demonstrated significant differentiation from one or more ethnic groups of Sri Lanka at two loci (DXS6789, DXS7424) out of the nine common loci compared.

These results suggest that allele distribution of the four ethnic groups of Sri Lanka is very similar to the two tested Indian populations and the Bangladesh population. Pakistani population also exhibit substantial similarity to Sri Lankan ethnic groups, though not to the same extent of Indian and Bangladesh populations. On the contrary, allelic distribution of many X-STR loci in Sri Lankan ethnic groups differ from Southeast Asian, East Asian, European and African populations. Among them, East Asian and African populations are the most genetically distant populations to Sri Lankans.

To further clarify the relationship between Sri Lankans and the world populations, pairwise Fst values were averaged over eight of the 16 studied X-STR loci (DXS10148, DXS10135, DXS8378, DXS7132, DXS10079, DXS10074, HPRTB and DXS7423) and were represented in a multidimensional scaling (MDS) plot (Fig. 2). The plot graphically illustrates the level of similarity between 18 world populations including the four Sri Lanka ethnicities (Sinhalese, SL Tamil, Indian Tamil, Moors, Bhil India[17], Bangladesh[19], Malaysia[22], Thailand[23], China[24], Japan[24], Taiwan[29], Germany[31], Italy[35], Sweden[40], Denmark[41], North Portugal[42], Somalia[41], and Ivory Coast[43]). According to the results observed, Sri Lankans were clustered together not only with Indians and Bangladeshi, but also with Europeans. Indian Tamils were placed towards the periphery of this main cluster, while Southeast Asians, East Asians and Africans were placed at a distant, outside the main cluster.

This presentation of MDS plot aligns well with the historical claims of population movements in Eurasia. Sinhalese are believed to be descended from Indo- Aryans, who set forth from boarders of Caspian and Black sea towards Europe and South Asia, early in the third millennium BC. Accordingly, many scholars hold the view that along with the Indo-Aryan language family, many Europeans and South Asian civilization of today share common genetic background reflecting their Bronze age common ancestors. Tamils on the other hand are believed to have descended from the indigenous people of Indian subcontinent. However, Sri Lankan Tamils have admixed with Sinhalese nearly over two millennia, unlike the Indian Tamils, which might explain their relative positions in the MDS plot.

Ld And Haplotype Analysis

Population studies of various countries have illustrated that the LD is population specific[41,45,46] and does not necessarily present between markers with close physical proximity[47,48]. In the present study, Sri Lankan Tamils did not exhibit LD within any of the four clusters, while Sinhalese displayed LD within three of the four studied clusters after adjusting for multiple comparisons (Table 14). Further, in cluster I and IV, LD was detected only among Sinhalese population, and only between a single pair of loci (DXS10135 & DXS8378 in cluster I and DXS7424 & DXS101 in cluster II). Among these markers, DXS10135 & DXS8378 of cluster I are separated by 13.1 kb. However, there was no LD detected between DXS10148 & DXS10135 (of the same cluster), which are located only 1 kb apart. On the other hand, the cluster II, which had the highest presentation of LD among the four clusters, displayed a highly significant (P = 0.0000) LD between two of its adjacent markers, DXS10074 and DXS10075 in three ethnicities (Sinhalese, Indian Tamils and Moors). In addition, LD was also detected between DXS7132 and DXS10075, the outermost two of the four markers of cluster II, among the Sinhalese (P = 0.0000). In contrast, in cluster III, none of the markers showed a significant LD (corrected P = 0.0167), although there was a marginal LD (P = 0.0174) detected between DXS6801 and DXS6809 for Sinhalese population.

Table 14
P values for pairwise linkage disequilibrium results for the four ethnicities.

Pair Locus	Sinhala	Sri Lankan Tamil	Indian Tamil	Moors
Cluster 1				
DXS10148,DXS10135	0.6029	0.0707	0.1486	0.4051
DXS10148,DXS8378	0.1437	0.8618	0.2928	0.1165
DXS10135,DXS8378	**0.0046**	0.6634	0.2217	0.2390
Cluster II				
DXS7132, DXS10079	0.5373	0.0378	0.5372	0.6837
DXS7132, DXS10074	0.0410	0.2924	0.1040	0.2472
DXS10079, DXS10074	0.1029	0.0661	0.0202	0.7661
DXS7132, DXS10075	**0.0000**	0.3472	0.2258	0.2234
DXS10079, DXS10075	0.7620	0.1582	0.0197	0.2328
DXS10074, DXS10075	**0.0000**	0.3612	**0.0000**	**0.0000**
Cluster III				
DXS6801,DXS6809	0.0174	0.8943	0.4792	0.0251
DXS6801, DXS6789	0.1241	0.6836	0.9615	0.0285
DXS6809, DXS6789	0.3075	0.5340	0.7291	0.9068
Cluster IV				
DXS7424, DXS101	**0.0159**	0.5544	0.4060	0.2435
DXS7424, DXS7133	0.2616	0.7064	0.5845	0.1790
DXS101, DXS7133	0.0680	0.0972	0.9476	0.5594
Cluster III and IV				
DXS6801,DXS7424	0.5476	0.7227	0.0815	0.0198
DXS6809,DXS7424	0.7421	0.2305	0.5808	0.1146
DXS6789,DXS7424	0.6440	0.8464	0.6726	0.1738
DXS6801,DXS101	0.9240	0.4147	0.8663	0.1879
DXS6809,DXS101	0.2092	0.9795	0.7262	0.0803
DXS6789,DXS101	0.3769	0.7291	0.3066	0.3447
DXS6801,DXS7133	0.8799	0.7682	0.3120	0.9418
DXS6809,DXS7133	0.1199	0.5175	0.7853	0.5001
DXS6789,DXS7133	0.2477	0.5693	0.2664	0.3527

Significant P values (after correcting for multiple comparisons) are indicated in bold font.

LD is generally expected to be high in populations, which are either small, reproductively isolated or having low population growth[47]. Indian Tamils who are tea estate workers, restricted mostly to central part of the country fits well into this description. In Sri Lanka, there are about 850,000 Indian Tamils, which had very limited genetic mixing with other populations due to social and cultural reasons. Nevertheless, only one loci pair among Indian Tamils showed LD according to our results. As pointed out by Kling et al[49], to acquire adequate power to detect LD in X-STR analysis, sample sizes over 200 are often needed. When considering the limited number of Indian Tamil samples used in the present study in calculating LD (63 males), it is possible that the low sample number to have masked the actual existence in LD among Indian Tamils. The same limitation is also valid for Sri Lankan Tamil and Moor ethnicities in which similarly low level of LD were detected. On the other hand, analysing a small sample from a large population can create LD, even among loci which are in linkage equilibrium, due to the under-representation of haplotypes[50]. In this light, it is advisable to increase the number of Sinhalese samples used for LD calculations (258 males in the present study), considering that there are about 15 million Sinhalese in the country. On the other hand, the high level of LD observed within Sinhalese might have resulted from its largely admixed nature. All these facts highlight the requirement of a more exhaustive data set with increased power to detect LD, before concluding on the true pattern of LD among Sri Lankan ethnicities.

In general, the LD reported earlier for South Asian and South East Asian populations like Bangladeshi[19], Indian Bhil tribe[17] and Malaysians[22] was quite low; i.e. no LD was reported among cluster I (DXS10148, DXS1035 and DXS8378) or in cluster II markers (DXS7132, DXS10079 and DXS10074). The marker, DXS10075, in which LD was detected with DXS7132 for Sinhalese was not investigated in any of these populations to make a comparison. However, it is noteworthy that all these studies were conducted with male samples between 100 and160, which might have posed a limitation in detecting true LD in these populations. On the other hand, LD was not observed for cluster III or cluster IV in a study that investigated 302 Pakistan males[20], the only Asian population for which LD data are available for these clusters. These data are suggestive of the existence of a complex pattern of LD among Asians. In contrast, a high level of LD was reported within these four clusters for Europeans like Swedish[51] and German[52–54] populations, in studies conducted with 450–800 male individuals.

The genetic stability of the linked clusters and the degree of dependence between them can have a major influence on most forensic and kinship applications. Therefore, in addition to the pairwise LD analysis within separate clusters, pairwise LD was also analyzed between markers belonging to cluster III and IV, due to their relative close physical proximity (6.78 cM) compared to the other clusters. However, LD was not detected between any of the marker pairs ($P > 0.05$ for eight out of nine marker pairs) as shown in Table 14 confirming the independent nature of the two clusters irrespective their physical location.

The haplotype frequencies obtained for the four ethnic groups for all four clusters are listed in Supplementary Tables S7-S10. In most of the previously published population studies, cluster 1V comprises only DXS7424-DXS101 without DXS7133[54–56]. Likewise, the four markers included in the cluster II had been analyzed in two different combinations; i.e. DXS7132-DXS10079-DXS10074 (Argus X-12 kit) and DXS10079-DXS10074-DXS10075[53]. Therefore, to allow comparisons with these previous studies, the haplotype frequencies of above combinations are also listed in Supplementary Tables S11-S14.

As shown in Supplementary Table S15, typing of male subsamples from the four ethnic groups yielded a total of 604, 246, 208 and 223 different haplotypes for Sinhalese, Sri Lankan Tamils, Indian Tamils and Moors, respectively. Cluster I produced the highest number of haplotypes for Sinhalese and Sri Lankan Tamils, while both cluster I and II produced the highest number of haplotypes for Indian Tamils and Moors. Among all the observed haplotypes, 96.85% of

Sinhalese, 82.11% of Sri Lankan Tamil, 83.17% of Indian Tamil and 85.65% of Moor haplotypes showed frequencies < 0.020. Moreover, the most common haplotype was observed at a frequency ≤ 0.065 in all the four ethnicities, indicating the suitability of the selected X-STR clusters for haplotype based kinship analysis.

Among the four clusters, both clusters I and II proved relatively more informative for all four ethnicities as reflected by the haplotype diversity (HD = 0.9964–0.9987 and 0.9935–0.9973 respectively). In cluster I, this observation may have caused by the two highly polymorphic markers, DXS10135 and DXS10148 as described above and is consistent with other previously published population data[17,19,24]. Cluster II, on the other hand carries four markers compared to the other clusters, which consists of three markers each. This increased number of markers may have generated higher haplotype diversity with respect to cluster II among the studied populations. On the contrary, cluster III and cluster IV are equally informative (HD = 0.9873–0.9908 and 0.9882–0.9923 respectively), though not to the same extent as the clusters I and II. Nevertheless, in general, all four clusters showed a high haplotype diversity for all four ethnicities

Conclusion

In this work, we report for the first time, X chromosome based population genetic data for all four major ethnic groups in Sri Lanka which covers 99.5% of the total population. According to our results, the 16 X-STR assay system used in the current study is highly polymorphic and exhibited high forensic efficiency for all four ethnicities tested indicating its suitability to be used in both evolutionary genetic analysis and forensic applications of Sri Lankans. The present study has also revealed subtle but statistically significant differences in X-STR based allele frequency distribution of Indian Tamils with Sinhalese and Moors, contrary to the highly homogeneous genetic outlook portrayed by autosomal STR analysis. While suggesting a sex biased demographic history for Sri Lankan ethnicities, this observation recommends the use of a separate X-STR allele frequency database for Indian Tamils for forensic and kinship application purposes. In contrast, the observed genetic admixture present within Sinhalese, Sri Lankan Tamil and Moor ethnicities suggest the possible use of a common allele database for the purpose. Further, the genetic distances observed among the Sri Lankans and other nationalities in the world visualized in the MDS plot render evidence to the ancient linguistic origin of Sri Lankan ethnicities- Indo-Aryan origin of Sinhalese and Dravidian origin of Tamil populations- which had been later affected to various degrees through genetic admixing between them. LD was detected along the X chromosome in all ethnic groups except Sri Lankan Tamils, which need to be considered during the likelihood calculations of kinship resolution and person identification. Further, the patterns of LD observed, which differ substantially among the four ethnicities request the use of different haplotype frequency databases for the four ethnicities for forensic purposes. Although the results of haplotype analysis suggest that the four studied X-STR clusters can provide a powerful tool for kinship testing and relationship identification of Sri Lankan ethnicities, a more exhaustive sampling of the two Tamil groups and Moors is recommended to confirm the LD and haplotype based observations.

Methods

Sample preparation and DNA extraction

The study was conducted with the approval of the Ethics Review Committee, Institute of Biology, Sri Lanka (ERC IOBSL 135 11 15) and the study was performed in line with the principles of the Declaration of Helsinki. Written informed consent was obtained from all individual participants included in the study. Finger pricked blood samples were collected from 838 unrelated individuals from the four ethnic groups in the Sri Lankan population; 426 samples from Sinhalese (60.6% males), 154 samples from the Sri Lankan Tamils (50% males), 128 samples from Indian Tamils (49.2% males), and 130 samples from Sri Lankan Moors (51.5% males). Genomic DNA was extracted using

Chelex-100 method[57] and subjected to PCR amplification using the single tube 16 X-STR multiplex system described in Perera et al[14]. Amplified products were resolved with capillary gel electrophoresis using ABI 3500 Genetic Analyzer (Thermo Fisher Scientific, USA) and data analysis allele designation was performed using GeneMapper IDX software (Thermo Fisher Scientific, USA).

Statistical analysis

All population genetic parameters were calculated using ARLEQUIN 3.5. Allele and haplotype frequencies, exact test of differentiation for male and female allele frequencies, conformity of female subsamples to Hardy-Weinberg equilibrium and pairwise test of LD between pairs of markers within clusters of linked loci in the male sub samples were analysed separately for all four ethnic groups. LD between the markers of cluster III and IV was also analysed, considering the close physical proximity of the two clusters (6.78 cM). Forensic parameters, i.e. polymorphism information content (PIC), expected heterozygosity (He), mean exclusion chance in deficiency (MEC_{Kruger}), mean exclusion chance in Duos ($MEC_{Desmaris\ duo}$), mean exclusion chance in trios ($MEC_{Desmaris\ trio}$), power of discrimination for females (PD_F), power of discrimination for males (PD_M) were calculated using chromosome X web (http://www.chrx-str.org) for all X-STR loci for all four ethnic groups. For those loci under LD, the parameters were reevaluated based on frequencies of haplotypes defined by them with MATLAB software (version R2017a).

Locus-by-locus analysis of molecular variance (AMOVA) was conducted by grouping the four ethnicities based on their linguistic origin; i.e. Indo-Aryan origin (Sinhalese) vs Dravidian origin (Tamils and Moors) to create a hierarchical structure. Nei's average number of pairwise differences within and between populations (Nei and Li, 1979) was used to detect pairwise differences among the four ethnic groups. The null distribution of pairwise Fst values was obtained by permuting haplotypes between populations for 1000 times. The significance level was kept at 0.05 and all analyses were conducted using Arlequin software v.3.5.1.2. Further, for easier visualization of the observed genetic distances, a phylogenetic tree was also constructed with MEGA6 software using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) method. The reliability of phylograms was estimated by bootstrapping 2000 replicates over loci and the extended majority rule consensus trees were inferred. To compare the Sri Lankan ethnicities with populations from other countries in the world, locus by locus pairwise genetic distances (Fst) were generated based on data available in the literature. Resulting Fst values were averaged over loci and represented in a multidimensional scaling (MDS) plot using SPSS v.15.0 statistical package.

Data Availability

The datasets generated during the current study are available from the corresponding author on reasonable request.

Declarations

Acknowledgements

The work was supported by National Research Council of Sri Lanka (Grant No: 16-023) and Genetech Research Institute, Colombo 08, Sri Lanka.

Author contributions

N. P. and G. G. designed the study. Material preparation, data collection and analysis were done by N.P. Preparation of the first draft of the manuscript was done by N P. and G.G. The manuscript was reviewed and edited by G.G. and G. R. All authors read and approved the final manuscript.

Competing interests: The authors declare no competing interests.

Conflicts of interest/Competing interests

The authors declare that they have no conflict of interest.

Data availability

The datasets generated during the current study are available from the corresponding author on reasonable request.

References

1 Census & Statistics. Population by ethnic group according to districts, Sri Lanka, http://www.statistics.gov.lk/PopHouSat/CPH2011/index.php?fileName=pop42&gp=Activities&tpl=3, (2012).

2 Seligman, C. & Seligman, B. *The Veddas*. (Cambridge University Press, 1911).

3 Kshatriya, G. K. Genetic Affinities of Sri Lankan Populations. *Human Biology***67**, 843-866 (1995).

4 Hussein, A. *The Lion and the Sword; An Ethnological Study of Sri Lanka*. (Printel Publishers (Pvt.) Ltd., 2001).

5 Dewaraja, L. *The Muslims of Sri Lanka: one thousand years of ethnic harmony 900-1915*. (The Lanka Islamic Foundation, 1994).

6 Illeperuma, R. J., Mohotti, S. N., De Silva, T. M., Fernandopulle, N. D. & Ratnasooriya, W. D. Genetic profile of 11 autosomal STR loci among the four major ethnic groups in Sri Lanka. *Forensic. Sci. Int. Genet.***3**, e105-106, doi:10.1016/j.fsigen.2008.10.002 (2009).

7 Illeperuma, R. J. *et al.* Haplotype data for 12 Y-chromosome STR loci of Sri Lankans. *Forensic. Sci. Int. Genet.***4**, e119-120, doi:10.1016/j.fsigen.2009.10.003 (2010).

8 Ranaweera, L. *et al.* Mitochondrial DNA history of Sri Lankan ethnic people: their relations within the island and with the Indian subcontinental populations. *J. Hum. Genet.***59**, 28-36, doi:10.1038/jhg.2013.112 (2014).

9 Chen, P. *et al.* Genetic structure and polymorphisms of Gelao ethnicity residing in southwest china revealed by X-chromosomal genetic markers. *Sci. Rep.***8**, 14585, doi:10.1038/s41598-018-32945-7 (2018).

10 Zhang, S. *et al.* Population genetic study of 34 X-Chromosome markers in 5 main ethnic groups of China. *Sci. Rep.***5**, 17711, doi:10.1038/srep17711 (2015).

11 Butler, J. M. *Advanced Topics in Forensic DNA Typing: Methodology*. (Elsevier Academic Press, 2011).

12 Szibor, R. *et al.* Use of X-linked markers for forensic purposes. *Int. J. Legal. Med.***117**, 67-74, doi:10.1007/s00414-002-0352-5 (2003).

13 Szibor, R. X-chromosomal markers: past, present and future. *Forensic. Sci. Int. Genet.***1**, 93-99, doi:10.1016/j.fsigen.2007.03.003 (2007).

14 Perera, N., Galhena, G. & Ranawaka, G. Development of a 16 X-STR multiplex PCR system for kinship analysis and its applicability for the Sinhalese population in Sri Lanka. *Int. J. Legal. Med.***135**, 161-166, doi:10.1007/s00414-020-02450-6 (2021).

15 Simmons, M. P. A confounding effect of missing data on character conflict in maximum likelihood and Bayesian MCMC phylogenetic analyses. *Mol. Phylogenet. Evol.***80**, 267-280, doi:10.1016/j.ympev.2014.08.021 (2014).

16 Ingley, S. J., Reina, R. G., Bermingham, E. & Johnson, J. B. Phylogenetic analyses provide insights into the historical biogeography and evolution of Brachyrhaphis fishes. *Mol. Phylogenet. Evol.***89**, 104-114, doi:https://doi.org/10.1016/j.ympev.2015.04.013 (2015).

17 Shrivastava, P., Jain, T., Gupta, U. & Trivedi, V. B. Genetic polymorphism study on 12 X STR loci of investigator Argus X STR kit in Bhil tribal population of Madhya Pradesh, India. *Leg. Med. (Tokyo)***17**, 214-217, doi:10.1016/j.legalmed.2014.11.004 (2015).

18 Mukerjee, S., Ghosh, T., Kalpana, D., Mukherjee, M. & Sharma, A. K. Genetic variation of 10 X chromosomal STR loci in Indian population. *Int. J. Legal. Med.***124**, 327-330, doi:10.1007/s00414-010-0456-2 (2010).

19 Sufian, A. *et al.* Genetic diversity study on 12 X-STR loci of investigator(R) Argus X STR kit in Bangladeshi population. *Int. J. Legal. Med.***131**, 963-965, doi:10.1007/s00414-016-1513-2 (2017).

20 Tariq, M. A., Sabir, M. F., Riazuddin, S. A. & Riazuddin, S. Haplotype analysis of two X-chromosome STR clusters in the Pakistani population. *Int. J. Legal. Med.***123**, 85-87, doi:10.1007/s00414-008-0294-7 (2009).

21 Tariq, M. A., Ullah, O., Riazuddin, S. A. & Riazuddin, S. Allele frequency distribution of 13 X-chromosomal STR loci in Pakistani population. *Int. J. Legal. Med.***122**, 525-528, doi:10.1007/s00414-008-0263-1 (2008).

22 Samejima, M., Nakamura, Y., Nambiar, P. & Minaguchi, K. Genetic study of 12 X-STRs in Malay population living in and around Kuala Lumpur using Investigator Argus X-12 kit. *Int. J. Legal. Med.***126**, 677-683, doi:10.1007/s00414-012-0705-7 (2012).

23 Vongpaisarnsin, K., Boonlert, A., Rasmeepaisarn, K. & Dangkao, P. Genetic variation study of 12 X chromosomal STR in central Thailand population. *Int. J. Legal. Med.***130**, 1497-1499, doi:10.1007/s00414-016-1363-y (2016).

24 Uchigasaki, S., Tie, J. & Takahashi, D. Genetic analysis of twelve X-chromosomal STRs in Japanese and Chinese populations. *Mol. Biol. Rep.***40**, 3193-3196, doi:10.1007/s11033-012-2394-1 (2013).

25 Liu, Q. L. *et al.* Development of multiplex PCR system with 15 X-STR loci and genetic analysis in three nationality populations from China. *Electrophoresis***33**, 1299-1305, doi:10.1002/elps.201100558 (2012).

26 Liu, Q. L. *et al.* Development of the nine X-STR loci typing system and genetic analysis in three nationality populations from China. *Int. J. Legal. Med.***125**, 51-58, doi:10.1007/s00414-010-0520-y (2011).

27 Nakamura, Y. & Minaguchi, K. Sixteen X-chromosomal STRs in two octaplex PCRs in Japanese population and development of 15-locus multiplex PCR system. *Int. J. Legal. Med.***124**, 405-414, doi:10.1007/s00414-010-0467-z (2010).

28 Samejima, M., Nakamura, Y. & Minaguchi, K. Population genetic study of six closely linked groups of X-STRs in a Japanese population. *Int. J. Legal. Med.***125**, 895-900, doi:10.1007/s00414-011-0618-x (2011).

29 Chen, M. Y., Ho, C. W., Pu, C. E. & Wu, F. C. Genetic polymorphisms of 12 X-chromosomal STR loci in Taiwanese individuals and likelihood ratio calculations applied to case studies of blood relationships. *Electrophoresis***35**, 1912-1920, doi:10.1002/elps.201300645 (2014).

30 Hwa, H. L. *et al.* Thirteen X-chromosomal short tandem repeat loci multiplex data from Taiwanese. *Int. J. Legal. Med.***123**, 263-269, doi:10.1007/s00414-008-0310-y (2009).

31 Edelmann, J., Lutz-Bonengel, S., Naue, J. & Hering, S. X-chromosomal haplotype frequencies of four linkage groups using the Investigator Argus X-12 Kit. *Forensic. Sci. Int. Genet.***6**, e24-34, doi:10.1016/j.fsigen.2011.01.001 (2012).

32 Edelmann, J. & Szibor, R. Validation of the X-linked STR DXS6801. *Forensic. Sci. Int.* **148**, 219-220, doi:10.1016/j.forsciint.2004.04.069 (2005).

33 Edelmann, J., Deichsel, D., Plate, I., Kaser, M. & Szibor, R. Validation of the X-chromosomal STR DXS6809. *Int. J. Legal. Med.***117**, 241-244, doi:10.1007/s00414-003-0369-4 (2003).

34 Edelman, J. *et al.* 16 X-chromosome STR loci frequency data from a German population. *Forensic. Sci. Int.***124**, 215-218, doi:10.1016/s0379-0738(01)00565-5 (2001).

35 Inturri, S., Menegon, S., Amoroso, A., Torre, C. & Robino, C. Linkage and linkage disequilibrium analysis of X-STRs in Italian families. *Forensic. Sci. Int. Genet.***5**, 152-154, doi:10.1016/j.fsigen.2010.10.012 (2011).

36 Robino, C., Giolitti, A., Gino, S. & Torre, C. Development of two multiplex PCR systems for the analysis of 12 X-chromosomal STR loci in a northwestern Italian population sample. *Int. J. Legal. Med.***120**, 315-318, doi:10.1007/s00414-006-0115-9 (2006).

37 Turrina, S. & De Leo, D. Population genetic comparisons of three X-chromosomal STRs (DXS7132, DXS7133 and GATA172D05) in North and South Italy. *Int. Congr. Ser.***1261**, 302-304, doi:10.1016/s0531-5131(03)01626-1 (2004).

38 Asmundo, A., Perri, F. & Sapienza, D. Allele distribution of two X-chromosomal STR loci in a population from Sicily (Southern Italy). *Int. Congr. Ser.***1288**, 346-348, doi:10.1016/j.ics.2005.11.133 (2006).

39 Massetti, S. *et al.* Analysis of 8 STR of the X-chromosome in two Italian regions (Umbria and Sardinia). *Forensic. Sci. Int. Genet. Suppl. Ser.***1**, 157-159, doi:10.1016/j.fsigss.2007.10.022 (2008).

40 Tillmar, A. O. Population genetic analysis of 12 X-STRs in Swedish population. *Forensic. Sci. Int. Genet.***6**, e80-81, doi:10.1016/j.fsigen.2011.07.008 (2012).

41 Tomas, C., Pereira, V. & Morling, N. Analysis of 12 X-STRs in Greenlanders, Danes and Somalis using Argus X-12. *Int. J. Legal. Med.***126**, 121-128, doi:10.1007/s00414-011-0609-y (2012).

42 Caine, L., Costa, S. & Pinheiro, M. F. Population data of 12 X-STR loci in a North of Portugal sample. *Int. J. Legal. Med.***127**, 63-64, doi:10.1007/s00414-012-0672-z (2013).

43 Pasino, S. *et al.* Allele and haplotype diversity of X-chromosomal STRs in Ivory Coast. *Int. J. Legal. Med.***125**, 749-752, doi:10.1007/s00414-011-0591-4 (2011).

44 Penna, L. S. *et al.* Development of two multiplex PCR systems for the analysis of 14 X-chromosomal STR loci in a southern Brazilian population sample. *Int. J. Legal. Med.***126**, 327-330, doi:10.1007/s00414-011-0629-7 (2012).

45 Tao, R. *et al.* Investigation of 12 X-STR loci in Mongolian and Eastern Han populations of China with comparison to other populations. *Sci. Rep.***8**, 4287, doi:10.1038/s41598-018-22665-3 (2018).

46 Yang, X. *et al.* Genetic analysis of 19 X chromosome STR loci for forensic purposes in four Chinese ethnic groups. *Sci. Rep.***7**, 42782, doi:10.1038/srep42782 (2017).

47 Ardlie, K. G., Kruglyak, L. & Seielstad, M. Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet.***3**, 299-309, doi:10.1038/nrg777 (2002).

48 Chakravarti, A. Population genetics--making sense out of sequence. *Nat. Genet.***21**, 56-60, doi:10.1038/4482 (1999).

49 Kling, D., Dell'Amico, B. & Tillmar, A. O. FamLinkX - implementation of a general model for likelihood computations for X-chromosomal marker data. *Forensic. Sci. Int. Genet.***17**, 1-7, doi:10.1016/j.fsigen.2015.02.007 (2015).

50 Slatkin, M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.***9**, 477-485, doi:10.1038/nrg2361 (2008).

51 Tillmar, A. O. *et al.* Analysis of linkage and linkage disequilibrium for eight X-STR markers. *Forensic. Sci. Int. Genet.***3**, 37-41, doi:10.1016/j.fsigen.2008.09.006 (2008).

52 Szibor, R. *et al.* Haplotyping of STR cluster DXS6801-DXS6809-DXS6789 on Xq21 provides a powerful tool for kinship testing. *Int. J. Legal. Med.***119**, 363-369, doi:10.1007/s00414-005-0550-z (2005).

53 Hering, S. *et al.* DXS10079, DXS10074 and DXS10075 are STRs located within a 280-kb region of Xq12 and provide stable haplotypes useful for complex kinship cases. *Int. J. Legal. Med.***120**, 337-345, doi:10.1007/s00414-005-0061-y (2006).

54 Edelmann, J., Hering, S., Kuhlisch, E. & Szibor, R. Validation of the STR DXS7424 and the linkage situation on the X-chromosome. *Forensic. Sci. Int.***125**, 217-222, doi:10.1016/s0379-0738(02)00005-1 (2002).

55 Yang, X. *et al.* Development of the 19 X-STR loci multiplex system and genetic analysis of a Zhejiang Han population in China. *Electrophoresis***37**, 2260-2272, doi:10.1002/elps.201500540 (2016).

56 Guo, J. *et al.* Genetic structure and forensic characterisation of 19 X-chromosomal STR loci in Guizhou Sui population. *Ann. Hum. Biol.***46**, 246-253, doi:10.1080/03014460.2019.1623911 (2019).

57 Walsh, P. S., Metzger, D. A. & Higuchi, R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques***10**, 506-513 (1991).

Figures

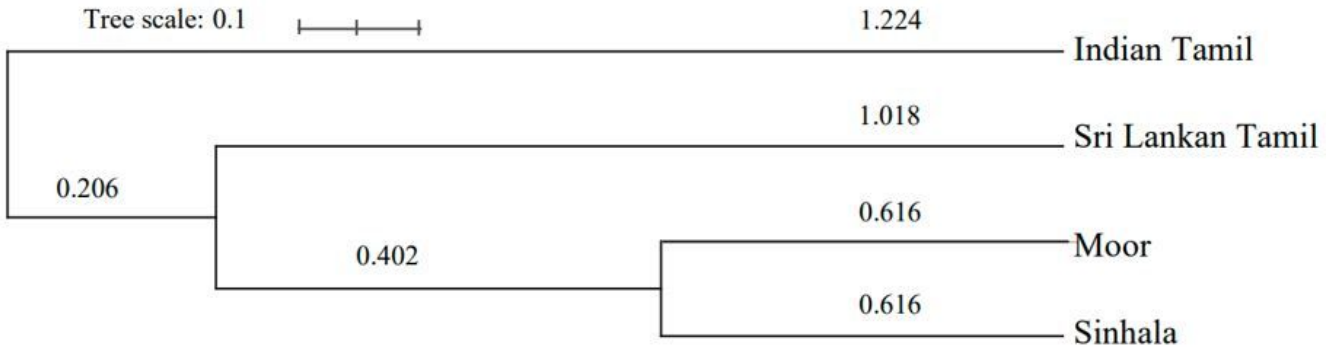

Figure 1

UPGMA phylogram for the four Sri Lankan ethnicities based on 16 X- STR data. The branch lengths are in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

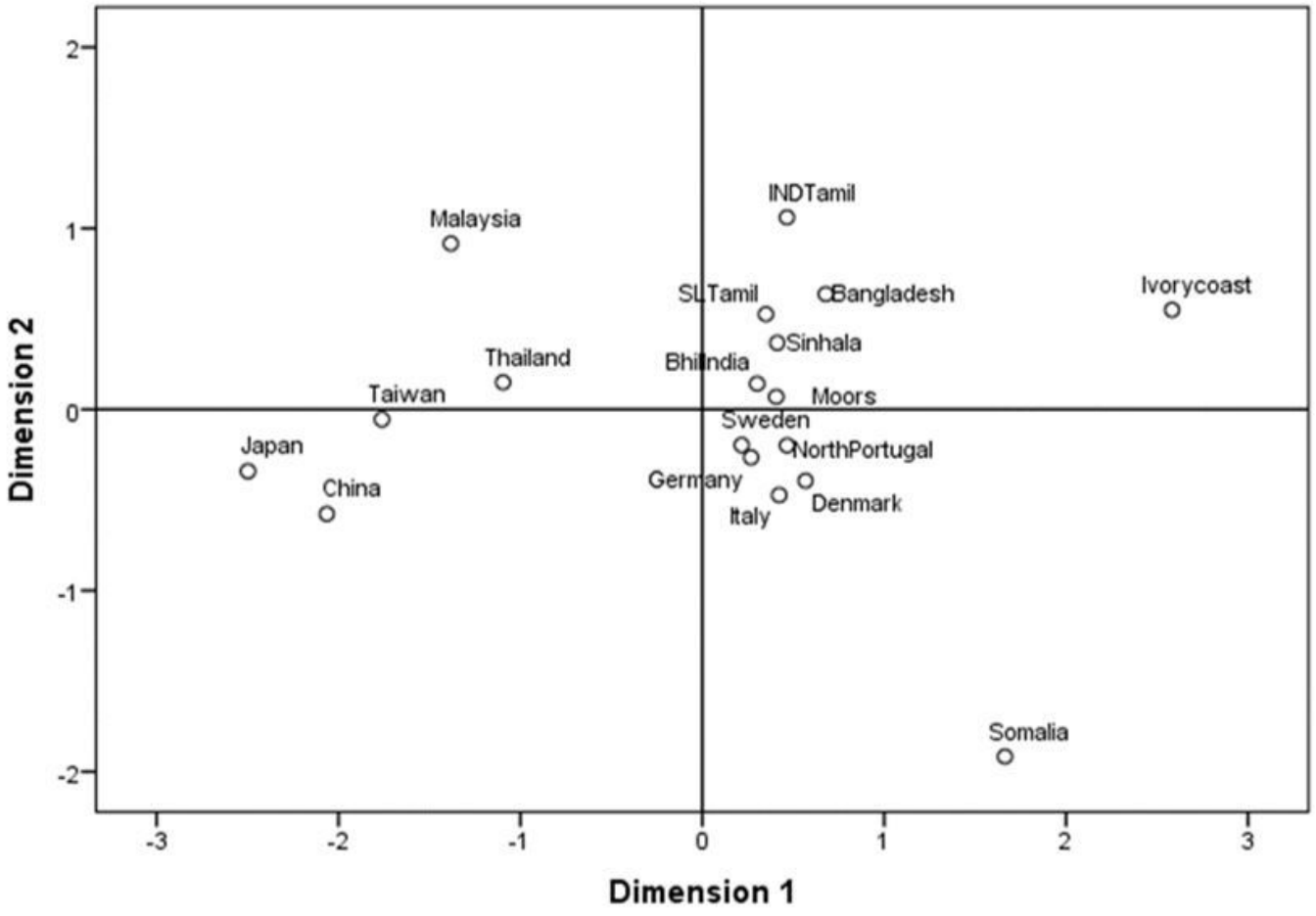

Figure 2

Two-dimensional MDS plot drawn from pairwise Fst values averaged over eight X-STR loci.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplimentaryTablesS1S15.pdf