Genetic analysis of novel Alu insertion polymorphisms in selected Indian populations.

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: LAYBOURN, S. ...et al., 2016. Genetic analysis of novel Alu insertion polymorphisms in selected Indian populations. American Journal of Human Biology, 28 (6), pp. 941-944.

Additional Information:

- This is the peer reviewed version of the following article: LAYBOURN, S. ...et al., 2016. Genetic analysis of novel Alu insertion polymorphisms in selected Indian populations. American Journal of Human Biology, 28 (6), pp. 941-944., which has been published in final form at http://dx.doi.org/10.1002/ajhb.22881. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Metadata Record: https://dspace.lboro.ac.uk/2134/21880

Version: Accepted for publication

Publisher: © Wiley

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
Genetic analysis of novel *Alu* insertion polymorphisms in selected Indian populations.

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>American Journal of Human Biology</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>AJHB-15-0178.R4</td>
</tr>
<tr>
<td>Wiley - Manuscript type:</td>
<td>Short Report</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
</tbody>
</table>
| Complete List of Authors:    | Laybourn, Susie; Loughborough University, School of Sport, Exercise and Health Sciences
                              | Akam, Elizabeth; Loughborough University, School of Sport, Exercise and Health Sciences
                              | Cox, Nick; Loughborough University, School of Sport, Exercise and Health Sciences
                              | Singh, Puneetpal; Punjabi University, Department of Human Genetics
                              | Mastana, Sarabjit; Loughborough University, School of Sport, Exercise and Health Sciences; |
| Keywords:                    | Alu insertions polymorphisms, India, Genetic variation, Tribes, Castes |
Genetic analysis of novel Alu insertion polymorphisms in selected Indian populations.

Susie Laybourn¹, Elizabeth Claire Akam¹, Nick Cox¹, Puneetpal Singh², Sarabjit S Mastana¹

1. Human Genomics Lab, School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, Leicestershire, LE11 3TU, United Kingdom. General telephone: +44 (0)1509 263171
2. Department of Human Genetics, Punjabi University, Patiala, Punjab, India.

• Number of text pages, plus bibliography, number of tables, figures, graphs, and charts

Short title - New Alu insertion elements in Indian Populations

Address correspondence to this author at the Human Genomics Lab: Dr Sarabjit Mastana, Loughborough University, Loughborough, Leicestershire, LE11 3TU, United Kingdom: Phone +44 (0)1509 223041; Fax: +44 (0)1509 226301; E mail s.s.mastana@lboro.ac.uk

Or

Dr Puneetpal Singh, Department of Human Genetics, Punjabi University, Patiala, India: Email singh.puneet@rediffmail.com

The authors thank all the participants for their cooperation and samples for this study. Financial support from SSEHS, Loughborough University is acknowledged.
Abstract

Objectives: Indian subpopulations (Chenchu, Koya and Lobana Sikh) were analysed at the genetic level for 12 Alu polymorphisms. These markers were then utilised to establish levels of genetic identity between the Indian populations and more widely between the Indian populations and a European population.

Methods: Previously collected blood samples were extracted using the phenol-chloroform method. The samples were utilised as templates for PCR using Alu specific primers and then analysed by agarose gel electrophoresis for the presence and absence of the approximately 300 bp insertion. Allele frequencies were calculated by the gene counting method and were tested for Hardy-Weinberg equilibrium, heterozygosities, inbreeding coefficient and G<sub>ST</sub> to assess the level of genetic differentiation.

Results: All of the Alu loci were polymorphic in the three Indian populations studied and their average observed heterozygosities ranged from 0.294 (Punjabi) to 0.357 (Koya). Allele and genotype frequency variation at 2b, 9a and ACE loci led to statistically significant pairwise differences among the three study populations. Overall population heterogeneity was observed for 7 out of 12 Alu polymorphisms.

Conclusion: The overall results show that these Indian samples, though displaying significant genetic variation and differences among themselves, form an Indian cluster, which as expected, is distinct from European sample (Russian). As Alus are easily analysed and quantified by standard and cost effective methodologies this further reinforces their utility as an effective population genetic marker.
Introduction

Alu insertion elements (Alus) represent the largest family of Short INterspersed Elements (SINEs) in humans and are named due to the presence of an AluI recognition site within the characteristic 300 bp unit. The human genome contains more than 1 million Alu repeats, which account for ~10% of the total nuclear DNA (Batzer et al 1996; Batzer and Deininger 2002; Stewart et al 2011). Alu repeats are organised into families and subfamilies, one of which is human specific (HS). HS Alus were inserted into the genome after the human divergence from a common ancestor with chimps thus these only appear in humans (Batzer et al. 1996), with insertions sites being commonly found in non-coding regions.

Alus have also been identified as highly stable due to their low insertion rate with only 100-200 new insertions occurring per millennia. These elements are also fixed due to lack of specific removal mechanisms in human genome (Batzer et al. 1996). This means individuals sharing Alu insertion polymorphisms have inherited from a common ancestor thus making the Alu insertion alleles identical by descent. So when using a number of Alus together a high forensic discrimination and phylogenetic analyses can be achieved (Mamedov at al., 2010) which is advantageous when analysing DNA for both ethnicity studies and forensic analyses.

India exhibits a large amount of genetic diversity for a country. Since the dispersal of humans from Africa, India has experienced large amounts of migration from many different populations resulting in a melange of ethnicities and cultural practices. In turn, this has led to both genetic divergences and convergences between subpopulations within India ((Majumder et al. 1999; Basu et al. 2003; Sahoo et al. 2006; Reich et al. 2009; Mastana, 2014; ArunKumar et al 2015).

While there are many reports on Alu polymorphisms in different populations of the world (Batzer et al. 1996), studies from the Indian subcontinent are limited to relatively small number of loci and populations (Majumder et al. 1999; Tripathi et al. 2008; Kshatriya et al. 2012; Saini et al 2012). In order to extend and document genetic variation of Alu polymorphisms, we analysed 8 new Alu insertion polymorphisms together with 4 extensively studied Alus (ACE, TPA, APO, PV92) (Majumder et al. 1999; Mamedov et al. 2010) in three distinct populations from India (1 caste population from Punjab (Lobana Sikhs) and 2 tribal populations (Chenchu and Koya) from Andhra Pradesh (now Telangana). The eight new Alu loci were selected from published paper (Mamedov et al. 2010) based on the observed heterozygosity levels (medium to high) and their adaptability to a multiplex reaction.

Methods

Subjects and study protocol

Genomic DNA samples were collected from three populations of India, Lobana Sikh caste population from Punjab region of Northern India, and two tribal populations (Chenchu and Koya) from Andhra Pradesh in South India (Mastana et al 1991; Mastana et al 2013, Papiha et al. 1997). The three populations represent groups who have both cultural and geographical barriers to intermarriage (Papiha et al. 1997). Brief details of the study populations are given below.
Chenchu tribe; a scheduled tribe, are also known as Chenchuvara inhabit three areas of Andhra Pradesh (AP); Mahabub Nagar, Kurnool and Guntur (Singh 1997; Papiha et al. 1997). They traditionally live in patriarchal and patrilineal nuclear families and have their own Chenchu language, which is part of the Dravidian family of languages (Bhasin et al. 1994, Bhasin and Walter 2000, Singh 1997). Primarily described as “shy hunter-gatherers” in 1694, they work on the land and in forests collecting produce and more recently have become agricultural labourers. In 1971, Chenchus numbered around 18,000 but only a minority still live in the traditional subsistence lifestyle (Singh 1997).

Koya tribe; also known as Konda Rajulu are inhabitants of Andhra Pradesh and have a larger population size (359,799, as of the 1981 census) (Singh 1997). They speak Gondi which has links with Dravidian languages common in Southern India (Singh 1997). Koya people are primarily farmers but exist in several endogamous groups which have their own occupations, for example blacksmiths brass workers and basket makers (Papiha et al. 1997). They live in villages in exogamous patrilocal clans and consanguinity between cousins along with uncles and nieces is common (Singh 1997).

Lobana Sikh is an endogamous now settled agricultural population of Punjab. In the recent past, they were a semi-nomadic tribe which traded in different commodities across India and abroad. Their population size is large (above 1 million) and they now profess to a range of occupations (Mastana 1989; Mastana et al. 1991; Mastana et al. 2013). Marriage patterns are strict and caste endogamy and clan exogamy is commonly practiced.

The Koya were sampled from Khamman District, and the Chenchu samples were collected from the Guntur and Hyderabad districts of Andhra Pradesh (Papiha et al. 1997). Lobana samples were collected from Patiala and Kapurthala districts of Punjab (Mastana et al. 1991).

All participants were healthy unrelated individuals who participated in our ongoing genetic research among Indian populations. The research was approved by the Loughborough University Ethical Advisory Committee and all participants gave full written consent. The DNA was extracted using the organic method (Sambrook et al. 1989).

**Genotyping method**

Alu polymorphisms were analysed by PCR amplification using specific primers as detailed by Mamedov et al., (2010). Individual allele genotypes were determined based on the respective DNA amplicon sizes visualised after gel electrophoresis (Mamedov et al. 2010; refer also S-Table 1 and S-Table 2).

**Statistical analysis**

Allele frequencies were calculated by the gene counting method and were tested for Hardy-Weinberg equilibrium. Chi-square test was used to compare frequencies between populations using PowerMarker Version 3.25 (http://statgen.ncsu.edu/powermarker/).

There is a lack of published studies on the 8 Alu loci employed in this study among Global, European and Indian populations, only a few studies are available from
Russia therefore comparative allele frequencies representing a European population (Russian) were collected from previously published literature (Mamedov et al. 2010 and Litvinov et al. 2008; Refer also S-Table 3). This population was considered as an outgroup for Indian populations.

Results

The sample size, insertion allele frequency, HWE p value, observed and expected heterozygosity and inbreeding coefficient (f) are given in Table 1. All Alu loci were polymorphic in three populations studied. Hardy-Weinberg equilibrium (HWE) deviation was observed for 8 populations/loci combinations (APOA1 in Chenchu, 2b, 8b, 12 and ACE in Punjabi, 9a, 8b and 12 in Koya). Pairwise genotype comparisons using chi-square showed that the Chenchu differed significantly from Koya at 2b, 9a,11, TPA and APO loci (all p values <0.05); and from Punjabis at 2a, 2b, 4a, 9a,11 and ACE loci, while Koya differed significantly from Punjabis at 2b, 9a,and ACE loci. Overall population heterogeneity was observed at 7 out of 12 loci (2a, 4a, 2b, 9a, 8b, 11 and ACE), indicating significant differences in genotype frequencies in the different populations. Punjabi population had the lowest observed heterozygosity (0.294) compared to both tribal groups (0.329 for Chenchu and 0.357 for Koya) for this set of Alu loci. Marked heterozygosity differences were observed for, 2b, ACE and APO different populations.  Inbreeding co-efficient (f) values also showed wide variation among the different populations (Table 1). The f values indicate that Punjabi population has substantial deficit of observed heterozygosity among different loci; 6 loci have heterozygosity below the average (0.294). Average GST value 0.033+ 0.032 suggest wide variation among loci and populations. Individual locus GST values varied from -0.004(Alu 12) to 0.107 (Alu 2b). A negative value of the GST was due to high similarity in Alu insertion allele frequency; while a high positive value for Alu 2b (0.107) and ACE (0.072) were due to substantial differences in allele frequencies of the study populations.

Eight loci in this study are new and have not been studied in European, global and Indian populations previously so it was not possible to compare these to infer genetic diversity. However, a Russian population data set (sample size 90-252 for different loci) was available (Mamedov et al. 2010 and Litvinov et al. 2008) and this was used for genetic distance calculations and correspondence analysis. We calculated Nei’s unbiased DA distance based on 12 loci and reduced the multidimensional matrix into a dendrogram by UPGMA method (Figure 1 and S-Table 3). Bootstrap values (out of 100) for this dendrogram are high giving confidence in use of Alu polymorphisms in understanding of population relationships.

In this figure, it is clear that there are distinct differences between the populations, Chenchu and Koya tribes belonging to South India form a close cluster which is joined by Lobana Sikhs and finally by Russians. Similar results were also obtained with correspondence analysis (not shown).

On developing the data set further, we assessed the use of these 12 loci for forensic and paternity purposes. The power of discrimination (for forensic calculation) was
high for all three populations analysed (>99.99%). Power of exclusion calculation for paternity testing showed that using these 12 loci one could exclude 42% (Lobana), 31% (Chenchu) and 24% (Koya) of the alleged fathers in a paternity investigation, which is typical for any two allele systems like SNPs and Alus.
**Discussion**

The analysis of \( Alu \) polymorphisms from three Indian populations highlights a level of genetic variation previously unreported with possible implications for the understanding of the evolution of genetic diversity among Indian populations. One should be cautious in interpreting the level and extent of genetic variation observed in these populations as the sample sizes are relatively small, although the sampling is homogeneous and well characterised. The results show that 8 new \( Alu \) loci are polymorphic in all populations with variable level of heterozygosity. There is paucity of data on new \( Alu \) polymorphisms to carry out comprehensive comparative analyses. Overall results of well-studied \( Alu \) loci like ACE, TPA, APO and PV92 are comparable to many studies from Indian subcontinent (Majumder et al. 1999; Veerraju et al. 2008; Tripathi et al. 2008; Kshatriya et al. 2012; Saini et al 2012). \( Alu \) elements provide useful amounts of variation in evolutionary heritage. Koya and Chenchu share the similar environmental and geographical location but differ significantly at a number of \( Alu \) loci. This genetic diversity between the two groups could be attributed to their different ethnic origins (Koya: Gond tribe and Chenchu: proto-Australoid). This inference is also supported by other genetic studies on the same South Indian populations (Papiha et al 1997). Papiha et al. (1997) further studied both tribal populations using a battery of blood groups, red cell enzymes and serum proteins. They found that the two tribes were significantly genetically different even though they have close geographical proximity. Veerraju et al. (2008) studied five \( Alu \) insertions (ACE, TPA, APO, PV92 and D1) in Chenchu and Koya tribes and found significant differences between these tribes. In their extended analysis, they also showed Chenchus clustered with caste populations of Uttar Pradesh and not with their geographical neighbours from Southern India. The Lobana Sikh population, as expected, shows significant differences from the tribal populations due to different geographical location and different genetic origins.

Analysis of four previously studied loci (ACE, TPA, APO, PV92) using DA genetic distances showed the distinct positions of tribal groups (Chenchu and Koya). In the dendrogram (not shown) based on these 4 loci only, Chenchu and Koya were closest to each other while Lobanas and Russian formed a separate cluster. One needs to be cautious about these relationships as these are based on very small number of loci and small sample size. Analysis of 12 loci, though better predictor of population relationships (based on bootstrap values) among the study populations, should also be interpreted with caution.

Overall our results demonstrate that these Indian samples, although displaying significant genetic variation and differences among themselves, still form an Indian cluster distinct from the European sample set (Russian) (Figure 1). These conclusions are strongly supported by other studies on STR, \( Alus \), mtDNA, Y chromosome and other expansive genome-wide evaluations of Indian populations (Basu et al. 2003; Sahoo et al 2006; Reich et al. 2009; Chaubey et al 2011; Mastana, 2014; ArunKumar et al 2015). Further populations studies are required on these \( Alus \) to deconvolute the level and extent of genetic variation and population origins.
Conclusion

Overall, results suggest that these Indian populations show significant allelic and genotypic variation which is patterned on geographical and ethnic origins.

Author Contribution

SL, SM, and EA analysed the data and drafted the manuscript. EA, SM, and SL designed the study, and directed implementation and data collection. SL, EA, and NC collected the data, and PPS provided necessary sample support. All authors read and agreed to this manuscript.

References


PowerMarker Version 3.25 (http://statgen.ncsu.edu/powermarker/).


Supporting Information

Supporting Information - Methods

[INSERT S-TABLE 1 HERE]

[INSERT S-TABLE 2 HERE]

[INSERT S-TABLE 3 HERE]
Table 1: Sample Size, Allele frequency, HWE, Heterozygosities, inbreeding coefficients and $G_{ST}$ statistics for 12 loci among 3 Indian populations

<table>
<thead>
<tr>
<th>Marker</th>
<th>Sample size</th>
<th>$f_{Alu}$</th>
<th>HWE $p$</th>
<th>$H_o$</th>
<th>$H_E$</th>
<th>$f$</th>
<th>$G_{ST}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chenchu</td>
<td>Koya</td>
<td>Lobana</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sikh</td>
<td>Sikh</td>
<td>Sikh</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chenchu</td>
<td>Koya</td>
<td>Lobana</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>64</td>
<td>56</td>
<td>62</td>
<td>0.695</td>
<td>0.527</td>
<td>0.468</td>
<td>0.225</td>
</tr>
<tr>
<td>4a</td>
<td>64</td>
<td>56</td>
<td>63</td>
<td>0.102</td>
<td>0.152</td>
<td>0.254</td>
<td>0.066</td>
</tr>
<tr>
<td>2b</td>
<td>65</td>
<td>56</td>
<td>60</td>
<td>0.423</td>
<td>0.607</td>
<td>0.206</td>
<td>0.747</td>
</tr>
<tr>
<td>9a</td>
<td>65</td>
<td>55</td>
<td>53</td>
<td>0.169</td>
<td>0.091</td>
<td>0.330</td>
<td>0.447</td>
</tr>
<tr>
<td>8b</td>
<td>65</td>
<td>54</td>
<td>53</td>
<td>0.677</td>
<td>0.722</td>
<td>0.481</td>
<td>0.903</td>
</tr>
<tr>
<td>10</td>
<td>65</td>
<td>56</td>
<td>60</td>
<td>0.077</td>
<td>0.134</td>
<td>0.117</td>
<td>0.502</td>
</tr>
<tr>
<td>11</td>
<td>65</td>
<td>53</td>
<td>52</td>
<td>0.808</td>
<td>0.679</td>
<td>0.606</td>
<td>0.634</td>
</tr>
<tr>
<td>12</td>
<td>65</td>
<td>54</td>
<td>54</td>
<td>0.138</td>
<td>0.167</td>
<td>0.157</td>
<td>0.798</td>
</tr>
<tr>
<td>ACE</td>
<td>65</td>
<td>56</td>
<td>63</td>
<td>0.731</td>
<td>0.625</td>
<td>0.413</td>
<td>0.417</td>
</tr>
<tr>
<td>TPA</td>
<td>65</td>
<td>56</td>
<td>63</td>
<td>0.431</td>
<td>0.589</td>
<td>0.476</td>
<td>0.591</td>
</tr>
<tr>
<td>APO</td>
<td>59</td>
<td>55</td>
<td>54</td>
<td>0.805</td>
<td>0.682</td>
<td>0.778</td>
<td>0.002</td>
</tr>
<tr>
<td>PV92</td>
<td>59</td>
<td>52</td>
<td>61</td>
<td>0.398</td>
<td>0.548</td>
<td>0.418</td>
<td>0.728</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td>0.330</td>
<td>0.357</td>
<td>0.294</td>
<td>0.348</td>
</tr>
</tbody>
</table>

$f_{Alu}$, Insertion allele frequency
$p$, Pearson chi square p value
$H_o$, Observed heterozygosity
$H_E$, Expected heterozygosity
$f$, Inbreeding coefficient
Figure 1: UPGMA dendrogram of 4 populations based on 12 Alu loci