Genetic variation of MHC Class I polymorphic *Alu* insertions (POALINs) in three subpopulations of the East Midlands, UK.

Sarabjit S. Mastana[1], Jasvinder S. Bhatti[2], Puneetpal Singh[3], Adam Wiles[1] and Jonathan Holland[1]

1. Human Genomics Lab., School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK.
2. Department of Biotechnology and Bioinformatics, Sri Guru Gobind Singh College, Sector 26, Chandigarh, India
3. Department of Human Genetics, Punjabi University, Patiala, Punjab, India.

Correspondence to this author at the Human Genomics Lab:

Dr Sarabjit Mastana,
Loughborough University,
School of Sport, Exercise and Health Sciences,
Loughborough, Leicestershire, LE11 3TU,
United Kingdom.
Phone +44 (0)1509 223041:
Fax : +44 (0)1509 226301;
E mail s.s.mastana@lboro.ac.uk

Word count: 3586 (including references and abstract)

ABSTRACT

Background: *Alu* elements are highly researched due to their useful nature as markers in the study of human population genetics. Recently discovered Major Histocompatibility Complex (MHC) polymorphic *Alu* insertions (POALINs) have not been examined extensively for genetic variation and their HLA associations.

Aims: The aim of this study is to assess the genetic variation between three populations using five recently discovered POALINs.

Methods and Subjects: The study examined 190 healthy, unrelated subjects from three different populations in the East Midlands (UK) for the presence or absence of five *Alu* elements (*Alu*HG, *Alu*MICB, *Alu*HJ, *Alu*TF and *Alu*HF) via the polymerase chain reaction followed by gel electrophoresis. Data were analysed for genetic variation and phylogenetic analyses.

Results: All *Alus* were polymorphic in study populations. Appreciable allele frequency variation was observed at number of loci. The British population was significantly different from both the Punjabi Jat Sikh and Gujarati Patel populations**,** though showing a closer genetic relationship to the Punjabi Jat Sikh population than the Gujarati Patel population (Nei's D_A = 0.0031 and 0.0064 respectively).

Conclusions: MHC POALINs are useful markers in the investigation of genetic variation and the assessment of population relationships and may have bearing on disease associations due to their linkage disequilibrium with HLA loci; this warrants further studies.

Keywords: POALIN; Alu; Major histocompatibility complex; East Midlands; Punjabi Jat Sikhs; Gujurati Patels.

Background

Alu insertion polymorphisms stem from the family of Short Interspersed Nuclear Elements (SINEs) that account for approximately 10% of the variation within human DNA with the human genome currently containing approximately 1.1 million *Alu* copies (Bennett et al. 2008; Stewart et al 2011). *Alu* sequences are derived ancestrally from the 7SL RNA gene and are mostly fixed in the human genome (Batzer and Deininger 1991). However, some *Alu* elements remain transcriptionally active, mobile and are not-fixed within the genome and these polymorphic *Alu* insertions (POALINs) are of interest for population genetic diversity, evolution, disease and forensic analysis. Benefits of studying POALINs as evolutionary markers are related to their unique mutational mechanism, specifically a lack of both back mutation and recurrent forward mutation (Deininger and Batzer, 1999; Batzer and Deininger, 2002). POALINs have proved useful in research with evidence reinforcing the out of Africa theory of human evolution (Stoneking et al. 1997; Antunez-de-Mayolo et al. 2002). York et al. (1999) showed that *Alu* insertion frequencies may contain greater phylogenetic information than allele frequencies derived from point mutations; further strengthening the usage of POALINs as a useful tool in the study of evolution.

Based upon commonly shared diagnostic mutations, genetic ages and sequence differences three fundamental categories of Alu sequences have been defined: *Alu*J, *Alu*S and *Alu*Y (Batzer et al. 1996; Kapitonov and Jurka 1996). Within the *Alu*Y sub group, *Alu* members such as *Alu*Ya5 and Yb8 have been shown to be useful candidates in the investigation of human ancestral haplotypes, population genetics and disease associations. *Alu*Yb8 appears to be human specific and has been linked to genetic diseases like acholinesteremia (Muratani et al. 1991) and Huntington's disease (Goldberg et al. 1993).

The major histocompatibility complex (MHC) is a highly polymorphic region found on human chromosome 6p21.31 and hosts a large number of genes

associated with immune function and disease. The main locus in this region is HLA (human leukocyte antigens) that can be divided into three sub-regions Class I, Class II and Class III. HLA region harbours a range of genetic polymorphisms like SNPs, (Walsh et al. 2003), microsatellite repeats (Malkki et al. 2005) and POALINs (Kulski et al. 2001, 2002; Dunn et al. 2002, 2003; Kulski and Dunn 2005; Yao et al. 2009; Kulski et al.2010; Shi et al 2014). The five POALINs recently discovered in the HLA class I region are *Alu*MICB, *Alu*TF, *Alu*HJ, *Alu*HG and *Alu*HF, and these are part of the *Alu*Yb8/*Alu*Ya5 subfamily of *Alus* (Kulski and Dunn, 2005). Yao et al. (2008) highlighted that the addition of the HLA-A locus to *Alu* haplotype analysis yields more information on the MHC haplotypes.

*Alu*HF, AluHG and A*lu*HJ are located within the alpha block of MHC molecule; AluHF is located slightly telomeric to the HLA-F gene, *Alu*HG is located between HLA-G and HLA-A and *Alu*HJ lies centromeric of HLA-J near the end of the alpha block. *Alu*TF lies in the intergenic area between the kappa and beta blocks of the region and *Alu*MICB lies within the first intron of the MICB gene closest to the centromere (see figure1)

[Insert Figure 1 here]

There are only few studies on HLA- POALINs and there are no studies on Indian or UK populations, this study aims to document the level and extent of genetic variation at 5 POALINS and assess their usefulness in population genetics and disease. Based on migratory nature of two study populations, it is hypothesized that the two Indian populations will have a significant genetic difference from the British population. This analysis would contribute significantly to assess genetic relatedness and provide a database of allele frequencies for future genetic, disease and forensic investigations.

METHODS

Subjects and study protocol

This study examined 190 apparently healthy, unrelated participants from three subpopulations of the East midlands (UK); one British population and two Indian migrant caste groups, Gujarati Patels (Hindus) and Punjabi Jat Sikhs.

A questionnaire was used to collect demographic information about their group membership and migration history. Participants were apparently healthy, unrelated to the grandparent generation and have been living in the East Midlands for 2-3 generations. All blood samples were collected with written consent and the study was approved by the Loughborough University Ethical Advisory Committee and followed the principles of the Declaration of Helsinki. The DNA was extracted from blood samples using the organic method (Sambrook et al. 1989). All samples were analysed without the knowledge of group membership.

Polymerase Chain Reaction (PCR) for detection of these POALINs have been described previously (Kulski et al, 2001, 2002a; Dunn et al, 2002, 2003b). The PCR was carried out using RedDy master Mix (Abgene) with variable MgCl2 concentrations (2mM for *Alu*HG, *Alu*MICB and *Alu*TF, and 3mM for *Alu*HJ and *Alu*HF). Specific master mixes for each POALIN were created based upon a standard ratio to provide a final reaction volume of 10µl. This basic ratio was 5µl RedDy Mix (with variable MgCl2), 3µl ultra-pure H_2O, 0.5µl forward primer, 0.5µl reverse primer. To this 1 µl of DNA of approximately 100ng/ µl was added. The DNA samples were amplified as per standard cycling conditions; initial denaturation of 5 minutes at 94°C followed by 35 cycles of denaturation at 94°C for 30 seconds, specific primer annealing at specific temperatures (see supporting information, table 1) for 30 seconds and extension at 72°C for 60 seconds with the final extension of 5 minutes at 72°C. 2-3% agarose gel electrophoresis was used for separation of alleles and genotypes were scored based on presence or absence of specific band sizes. The individual band sizes along with selected gel pictures are given in supporting information (Table 1 and Figure 1).

Statistical Analysis

Allele frequencies were calculated by the allele counting method. Hardy-Weinberg equilibrium was calculated for loci using observed and expected genotype frequencies. Chi-square test was used to compare genotype frequencies between populations. Haplotype analysis was performed using

Arlequin® software (version 3.11) (Excoffier, Laval, & Schneider, 2005). Genetic distances were calculated using Nei's D_A genetic distance (Nei et al. 1983) and resultant dendrograms were constructed using UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method (Sokal and Michener, 1958) using NTSYSpc programme (Rohlf, 2009). Multi-dimensional scaling analysis was also carried out using the NTSYSpc programme.

RESULTS

The sample size, genotype and allele frequencies with standard error, HWE chi-square, observed and expected heterozygosity and inbreeding coefficient (*f*) are given in Table 1. All *Alu* loci were polymorphic in three populations studied though no homozygote insertion genotypes were for observed for *Alu*MICB locus in any of the study populations. Significant departure ($p<0.05$) from Hardy-Weinberg equilibrium (HWE) was observed for 5 populations/loci combinations (*Alu*MICB in all populations and *Alu*TF and *Alu*HF in Gujarati Patels). Departures from HWE were not observed after using the Bonferroni correction (corrected p value = 0.003 for 15 chi-square tests). None of the HWE p-values were below this revised threshold.

Substantial allelic variation is apparent in these samples with average higher insertion frequency among British (0.272) followed by Punjabi Jat Sikh (0.217) and Gujarati Patels (0.202). At individual loci there is appreciable variation at *Alu*HG and *Alu*TF loci among the study populations (Table 1)

[Insert table 1 here]

Inbreeding co-efficient (*f*) values showed wide variation among the different populations and genetic loci (Table 1). Gujarati Patel population had the lowest observed heterozygosity (0.309) and highest inbreeding coefficient (0.029), suggesting some barriers to mating and endogamous structure. Punjabi Jat Sikh population inbreeding coefficient was -0.051 indicating heterogeneous (possibly multiple migrations) nature of the sample.

Pairwise genotype comparisons using chi-square showed that the Gujarati Patel population differed significantly at *Alu*HG from White British population (chi-square10.05, DF 2, $p<0.01$). Appreciable differences in allele frequencies at this locus also led to overall heterogeneity (chi-square 11.78, DF 4, $p<0.05$). All other pairwise comparisons were non-significant.

Allele frequency data was used to calculate Nei's D_A genetic distance (Nei et al. 1983) where it becomes apparent that the largest genetic distance is between the British and Gujarat Patel populations (0.0064) followed by British and Punjabi Jat Sikhs (0.0031) and the closest relationship was between Punjabi Jat Sikhs and Gujarati Patels (0.0012). The UPGMA dendrogram (was constructed for these populations (not shown here, as combined results are given in figure 2) which shows a distinct separation between the British and the two Indian populations and highlights the close relationship between the two Indian populations.

The haplotypes constructed from the all loci showed significant variation among populations. Punjabi population had highest number of observed haplotypes (30), British (29) and Gujarati the lowest (23) (Table 2). The haplotype frequencies showed that the absence of an *Alu* insertion at each locus was the most common haplotype with frequencies of 0.1989, 0.2416 and 0.3621 in the British, Punjabi Jat Sikh and Gujarati Patel populations respectively. All insertion haplotype was only estimated in British population and was missing in both Indian populations. Haplotype based dendrogram also illustrated similar typology as observed with allele frequencies.

[insert table 2 here]

Using published data from selected populations (Table 3), we worked out Nei's D_A genetic distances of 14 populations including 3 current study populations; the results of the UPGMA dendrogram are presented in figure 2. In this analysis one can see that population relationships are clearly reflecting the genetic origins and geographical similarities even when only five Alu loci

are used. Chinese sample from Malaysia had some very extreme insertion frequencies which led its isolation from other Chinese populations. Similar results were observed in multi-dimensional scaling (MDS) analysis plot (Figure 3). One should be cautious in interpreting relations in this analysis as achieved stress level is 0.145 which is considered as good to fair fit in terms of goodness of fit (Kruskal 1964). In MDS and dendrogram analyses, both Indian populations show a very close relationship. Similar results have been observed for other genetic loci (SNPs, STRs, *Alus*) among these populations (Ghelani et al. 2011; Mastana 2014).

[insert table 3 and figure 2 and 3 here]

COMMENT

This study was carried out to determine genetic variation between three populations as assessed by five POALINs as a primary objective and secondly to increase data available from previously unstudied groups. As such this study confirms the presence of genetic variation at HLA-*Alu* loci among British and Indian populations. It also documents significant allele frequency differences among the study populations which may have bearing on their association with HLA loci involved in disease and transplantation analysis. Overall our results are consistent with previous studies (Dunn et al. 2005, Kulski and Dunn 2005; Yao et al., 2008; 2009; Kulski et al 2011). British frequencies are comparable to Australians. There are no studies from India for comparison but overall range is similar to previous European and Asian studies (see table 3).

Watkins et al. (2003) documented that that the average *Alu* insertion frequencies for European and Indian populations are 0.559 and 0.544 respectively. Within our data the average insertion frequencies (for five loci) were 0.272, 0.217 and 0.202 for the British, Punjabi Jat Sikh and Gujurati Patel populations respectively. These frequencies are lower than those

observed in above large study confirming that HLA Alus are relatively young and recent insertion events derived from the Yb8 subfamily (Tian et al. 2008).

Based upon these observations and previous research on the MHC POALINs it may be useful to use these MHC POALINs in population studies. Genetic distance analysis clearly shows that POALINS have the potential for separating the populations even when only a limited number of loci in disease linked region are used. Allele frequency and haplotype based dendrograms displayed the same typology for the study populations.

Haplotype frequencies are as would be expected and are comparable to previous results (Kulski and Dunn, 2005; Shi et al 2014), showing a reduction in frequency as the number of *Alu* insertions are increased within the haplotype. Batzer et al. (1990) estimated that the probability of two or more independent insertions at the same nucleotide site was close to zero based on the knowledge that only 100-200 Alu elements in the human genome would be polymorphic after a million years. Therefore when a number of multiple POALIN haplotypes were identified, it could be considered as outcome of recombination within the MHC. Dunn et al. (2005) examined this and suggested that when more single POALIN haplotypes than multiple POALIN haplotypes were observed compared to expected values it supports the idea of recombination. They explain that due to the extremely low likelihood of multiple *Alu* insertions occurring at different loci within the same individual, it is much more probable that haplotypes with multiple POALINs have arisen through recombination of haplotypes with single but different *Alu* insertions. This is an interesting premise which warrants further analysis using HLA genes and POLAINs in these populations as a range of immune system diseases show variable prevalences in these groups.

The major limitation of this study is relatively small sample size, and absence of HWE at *Alu*MICB locus. At this locus no homozygote insertion genotype was observed in any of the study populations. Kulski et al. (2002) found that the AluyMICB insertion “occurred at relatively low gene frequency (0.113-0.118)” and that in their study populations only four people from 200

Northeastern Thai's and two people in 109 Western Australians were homozygous for the insertion. So it would be very optimistic to find a homozygous individual in small sample set of this study.

Based upon the results found in this study and the previous research (Dunn et al. 2005; Kulski and Dunn, 2005; Yao et al. 2009; Kulski et al 2011; Shi et al. 2014) it has been demonstrated clearly that the MHC POALINs are useful genetic markers for the study of human populations and diseases but to be truly helpful they must be studied in a greater number of populations around the world and their associations should be examined with reference to HLA loci and diseases.

Acknowldgements: Authors thank all participants for their participation in this study and Loughborough University for financial support.

References

Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P, Papiha SS, Hammer M, Yunis JJ, Damodaran C, Martinez de Pancorbo M, Caeiro JL, Puzyrev VP et al. 2002. Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. Electrophoresis 23: 3346-3356

Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hoppens CL, Deininger PL. 1990. Structure and variability of recently inserted Alu family members. Nucleic Acids Res 18(23): 6793-6798

Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJ, Deininger PL, Stoneking M.1996. Genetic variation of recent Alu insertions in human populations. J Mol Evol 42:22-29**.**

Batzer MA, Deininger PL. 1991. A human-specific subfamily of Alu sequences. Genomics 9: 481-487

Batzer MA, Deininger PL.2002. Alu repeats and human genomic diversity. Nat Rev Genet 3:370-379.

Bennett EA, Keller H, Mills RE, Schmidt S, Moran JV, Weichenrieder O, Devine SE. 2008. Active Alu retrotransposons in the human genome. Genome Res 18(12): 1875-1883

Deininger PL, Batzer MA. 1999. Alu repeats and human disease. Mol Genet Metab. 67: 183-193

Dunn DS, Choy MK, Phipps ME, Kulski JK.2007.The distribution of major histocompatibility complex class I polymorphic Alu insertions and their associations with HLA alleles in a Chinese population from Malaysia. Tissue Antigens. 70:136-143.

Dunn DS, Naruse T, Inoko H, Kulski JK. 2002. The association between HLA-A alleles and young Alu dimorphisms near the HLA-J, -H, and -F genes in workshop cell lines and Japanese and Australian populations. J Mol Evol 55: 718-726

Dunn DS, Romphruk AV, Leelayuwat C, Bellgard, Kulski JK. 2005. Polymorphic Alu insertions and their Associations with MHC Class I Alleles and Haplotypes in the Northeastern Thais. Ann Hum Genet 69: 364-372

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50.

Ghelani AM, Samanta A, Jones AC, Mastana SS.2011,Association analysis of TNFR2, VDR, A2M, GSTT1, GSTM1, and ACE genes with rheumatoid arthritis in South Asians and Caucasians of East Midlands in the United Kingdom. Rheumatol Int. 31:1355-1361.

Goldberg YP, Rommens JM, Andrew SE, Hutchinson GB, Lin B, Theilmann J, Graham R, Glaves ML, Starr E, McDonald H et al. 1993. Identification of an Alu retrotransposition event in close proximity to a strong candidate gene for Huntington's disease. Nature 362: 370-373

Kapitonov V, Jurka J. 1996. The age of Alu subfamilies. J Mol Evol 42: 59-65
Kruskal JB, 1964. Multidimensional Scaling by Optimizing Goodness-of-Fit to a Nonmetric Hypothesis. Psychometrika, 29:1–28.

Kulski JK, Dunn DS. 2005. Polymorphic Alu insertions within the Major Histocompatibility Complex class I genomic region: A brief review. Cytogenet Genome Res 110: 193-202.

Kulski JK, Martinez P, Longman-Jacobsen N, Wang W, Williamson J, Dawkins RL, Shiina T, Naruse T, Inoko H. 2001. The association between HLA-A alleles and an Alu dimorphism near HLA-G. J Mol Evol 53: 114-123.

Kulski JK, Dunn DS, Hui J, Martinez P, Romphruk AV, Leelayuwat C, Tay GK, Oka A, Inoko H. 2002. Alu polymorphism within the MICB gene and association with HLA-B alleles. Immunogenetics 53: 975-979.

Kulski JK, Shigenari A, Inoko H. 2011.Genetic variation and hitchhiking between structurally polymorphic Alu insertions and HLA-A, -B, and -C alleles and other retroelements within the MHC class I region.
Tissue Antigens. 78:359-377

Malkki M, Single R, Carrington M, Thomson G, Petersdorf E. 2005. MHC microsatellite diversity and linkage disequilibrium among common HLA-A, HLA-B, DRB1 haplotypes: Implications for unrelated donor hematopoietic transplantation and disease association studies. Tissue Antigens 66: 114-124
Mastana SS.2014.Unity in diversity: an overview of the genomic anthropology of India. Ann Hum Biol. 41:287-299

Muratani K, Hada T, Yamamoto Y, Kaneko T, Shigeto Y, Ohue T, Furuyama J, Higashino K. 1991. Inactivation of the cholinesterase gene by Alu insertion: Possible mechanism for human gene transposition. Proc Natl Acad Sci 88: 11315-11319

Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J. Mol. Evol. 19:153-170.

Rohlf FJ. 2009. NTSYSpc: numerical taxonomy system. ver. 2.21c. Exeter Software: Setauket: New York.

Shi L, Kulski JK, Zhang H, Dong Z, Cao D, Zhou J, Yu J, Yao Y, Shi L.2014

Association and differentiation of MHC class I and II polymorphic Alu insertions and HLA-A, -B, -C and -DRB1 alleles in the Chinese Han population. Mol Genet Genomics. 289:93-101

Sokal R, Michener C. 1958. A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin. 38: 1409–1438.

Stewart C, Kural D, Strömberg MP, Walker JA, Konkel MK, Stütz AM, Urban AE, Grubert F, Lam HY, Lee WP, Busby M, Indap AR, Garrison E, Huff C, Xing J, Snyder MP, Jorde LB, Batzer MA, Korbel JO, Marth GT; 1000 Genomes Project. 2011. A comprehensive map of mobile element insertion polymorphisms in humans. PLoS Genet. 7(8):e1002236

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA.1997.Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res. 7:1061-1071.

Walsh EC, Mather KA, Schaffer SF, Farwell L, Daly, Patterson N, Cullen M, Carrington M, Bugawan TL, Erlich H, Campbell J, Barrett J, Miller K, Thomson G, Lander ES, Rioux JD. 2003. An integrated haplotype map of the human major histocompatibility complex. Am J Hum Genet 73: 580-590.

Watkins WS, Rogers AR, Ostler CT, Woodings S, Bamshad MJ, Brassington A-M, Carroll ML, Nguyen SV, Walker JA, Prasad BVR, Reddy PG, Das PK, Batzer MA, Jorde LB. 2003. Genetic variation among world populations: Inferences from 100 Alu insertion polymorphisms. Genome Res 13: 1607-1618

Yao Y, Shi L, Shi L, Kulski JK, Chen J, Liu S, Yu L, Lin K, Huang X, Tao Y, Tokunaga K, Chu J. 2010.The association and differentiation of MHC class I polymorphic Alu insertions and HLA-B/Cw alleles in seven Chinese populations.Tissue Antigens.76:194-207.

Yao Y, Shi L, Shi L, Lin K, Tao Y, Yu L, Sun H, Huang X, Li Y, Chu J. 2009. Polymorphic Alu insertions and their associations with MHC class I alleles and haplotypes in Han and Jinuo populations in Yunnan Province, southwest of China. J Genet Genomics 36(1): 51-58

York DS, Blum VM, Low JA, Rowold DJ, Puzyrev V, Saliukov V, Odinokova O, Herrera RJ. 1999. Phylogenetic signals from point mutations and polymorphic Alu insertions. Genetica 107: 163-170

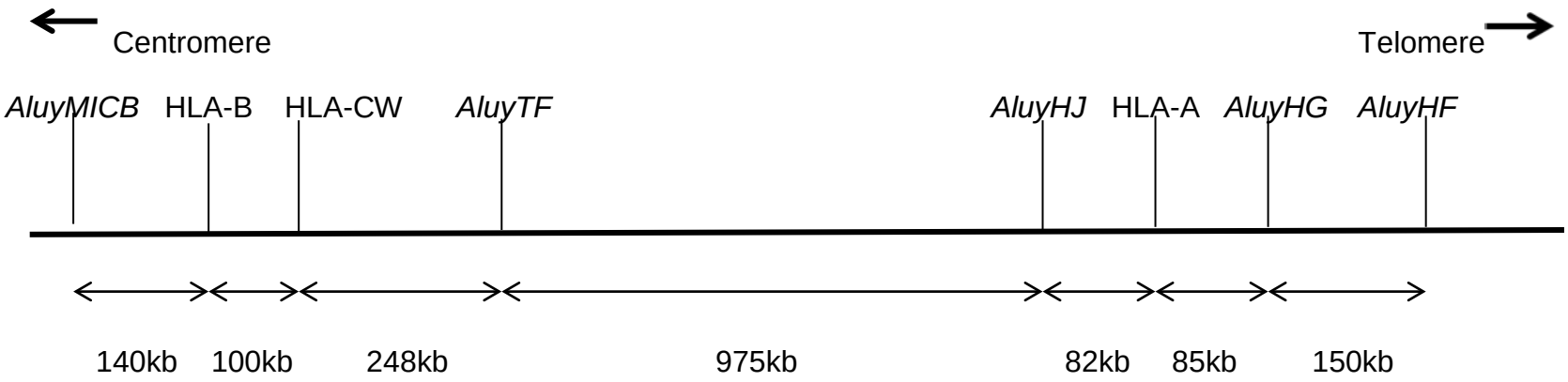
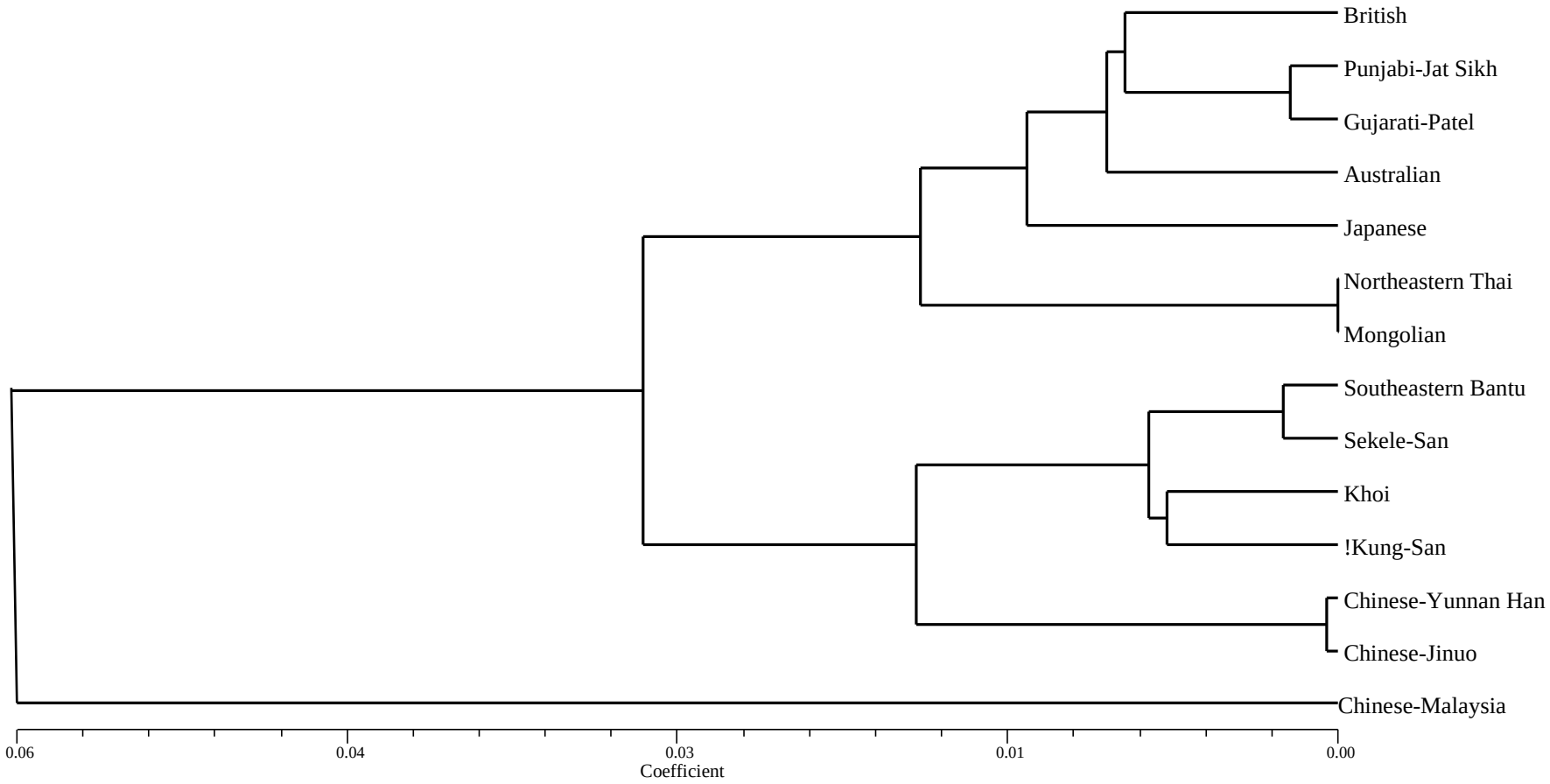


Figure 1. A Schematic representation of the five MHC POALINs found within the MHC class I region. (Adapted from Kulski and Dunn (2005))

Figure 2. Phylogenetic relationship of selected World populations based on 5 HLA-POALINS.



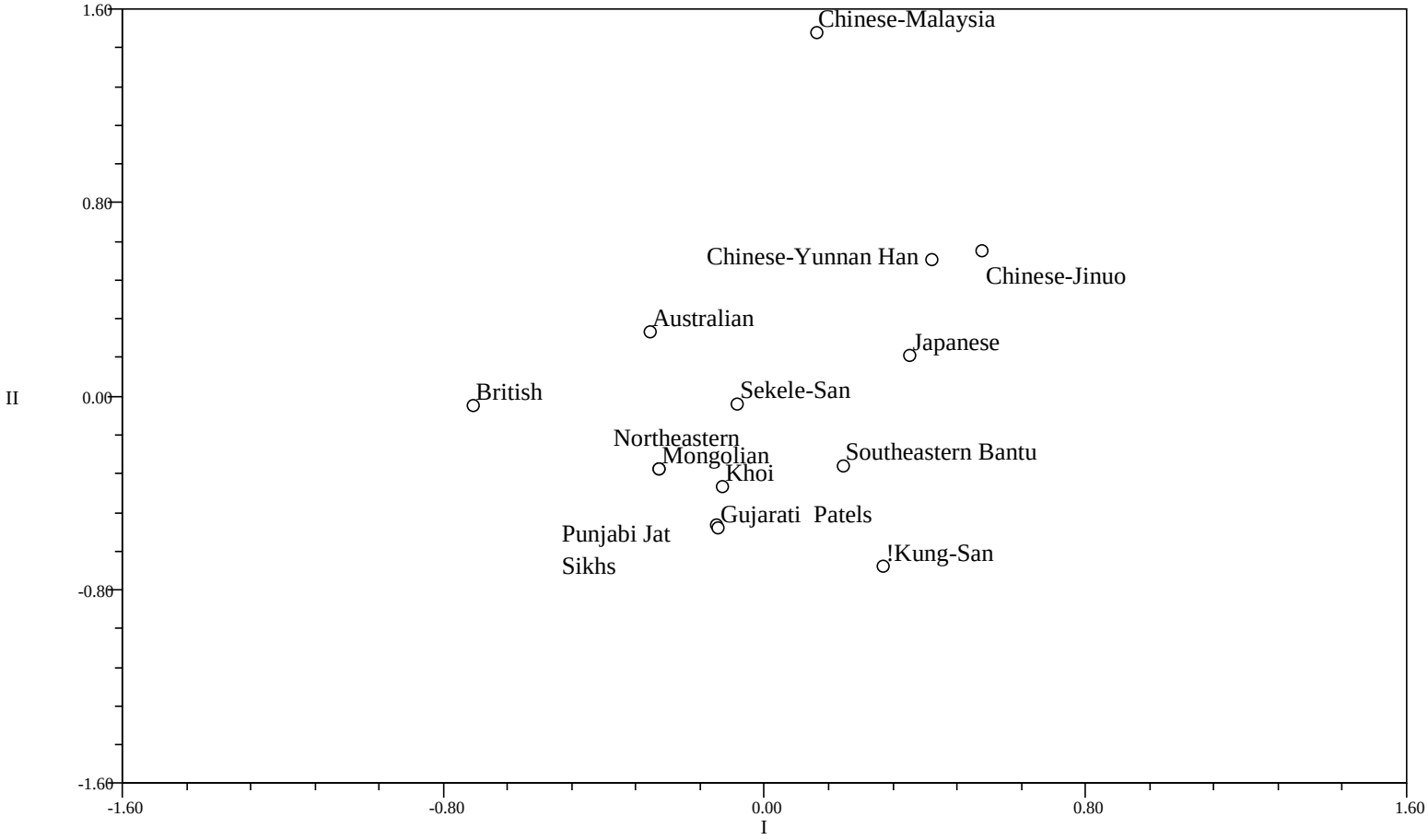


Figure 3. Multi-dimensional scaling plot of selected populations based on 5 POALINs. (Stress Value 0.145)

Table 1: Genotype and allele frequency distribution of 5 POALINs in 3 subpopulations of the East Midlands.

(Alu*D = Absence of Alu insertion (allele 1) & Alu*I = Presence of Alu insertion (allele 2))

Population/ POALINS	Genotypes (n)			Allele Frequencies		SE	Chi-square	Observed Heterozygosity	Gene Diversity	Inbreeding Coefficient
	II	ID	DD	Alu*I	Alu*D					
British (n= 61)										
AluHG	6	25	29	0.308	0.692	0.043	0.032	0.417	0.426	0.032
AluMICB	0	33	28	0.271	0.729	0.032	6.630*	0.541	0.395	-0.364
AluHJ	7	26	28	0.328	0.672	0.043	0.061	0.426	0.441	0.041
AluTF	5	19	37	0.238	0.762	0.039	1.204	0.317	0.362	0.126
AluHF	5	16	40	0.213	0.787	0.041	2.898	0.262	0.335	0.226
Average								0.355	0.370	0.012
Punjabi Jat Sikh (n = 66)										
AluHG	2	20	44	0.182	0.818	0.033	0.226	0.303	0.298	-0.011
AluMICB	0	27	38	0.208	0.792	0.031	4.467*	0.415	0.329	-0.254
AluHJ	8	30	28	0.348	0.652	0.041	<0.001	0.454	0.454	0.007
AluTF	2	15	49	0.144	0.856	0.032	0.399	0.227	0.246	0.085
AluHF	2	23	41	0.204	0.796	0.033	0.332	0.348	0.325	-0.063
Average								0.349	0.331	-0.051
Gujurati Patels (n=63)										
AluHG	1	15	46	0.137	0.863	0.031	0.031	0.242	0.237	-0.014
AluMICB	0	26	37	0.207	0.794	0.031	4.258*	0.412	0.327	-0.252
AluHJ	5	29	29	0.309	0.691	0.040	0.373	0.460	0.427	-0.069
AluTF	5	14	44	0.191	0.809	0.040	4.918*	0.222	0.308	0.286
AluHF	4	13	46	0.167	0.833	0.037	4.166*	0.206	0.277	0.264
Average								0.309	0.315	0.029

- * All HWE departures are significant at $p < 0.05$, After Bonferroni correction ($p < 0.003$), All loci are in HWE.

Table 2: Haplotype distribution of 5 HLA Alu loci among 3 subpopulations of the East Midlands (Order of Alus in haplotype is AluyHG, AluyMICB, AluyHJ, AluyTF and AluyHF. 1 refers to Deletion allele and 2 refers to insertion allele)

	British		Punjabi Jat Sikhs		Gujarati Patels	
Haplotype	**Frequency**	**SD**	**Frequency**	**SD**	**Frequency**	**SD**
1-1-1-1-1	0.1989	0.0611	0.2416	0.0944	0.3621	0.0598
1-1-2-1-1	0.1819	0.0511	0.1808	0.0607	0.0979	0.0415
2-1-1-1-1	0.1216	0.0500	0.1046	0.0477	0.0638	0.0467
1-2-2-1-1	0.0420	0.0288	0.0634	0.0334	0.0631	0.0336
1-1-1-1-2	0.0576	0.0375	0.0765	0.0442	0.0711	0.0275
1-2-1-1-1	0.0612	0.0348	0.0825	0.0464	0.0480	0.0290
2-1-1-1-2	0.0383	0.0279	0.0236	0.0263	0.0097	0.0101
2-2-1-1-1	0.0456	0.0320	0.0017	0.0054	0.0016	0.0043
1-1-2-1-2	0.0192	0.0266	0.0365	0.0325	0.0278	0.0248
2-1-1-2-1	0.0281	0.0257	0.0196	0.0233	0.0084	0.0080
1-1-2-2-1	0.0047	0.0123	0.0266	0.0270	0.0418	0.0258
1-2-1-1-2	0.0296	0.0238	0.0211	0.0182		
1-2-2-2-1	0.0245	0.0236	0.0043	0.0094	0.0282	0.0260
1-1-1-2-1	0.0347	0.0251	0.0363	0.0381	0.0456	0.0246
1-2-1-2-1	0.0098	0.0098	0.0160	0.0203	0.0311	0.0202
2-1-2-1-1	0.0192	0.0143	0.0137	0.0136	0.0279	0.0240
1-1-1-2-2	0.0137	0.0122	0.0083	0.0133	0.0326	0.0228
2-2-2-1-2	0.0092	0.0118			0.0044	0.0069
1-2-1-2-2	0.0091	0.0079	0.0041	0.0083		
1-1-2-2-2			0.0108	0.0150	0.0009	0.0036
2-2-1-2-1	0.0097	0.0161	0.0023	0.0055	0.0034	0.0062
2-2-1-1-1			0.0028	0.0070	0.0127	0.0152
2-2-2-2-1	0.0075	0.0142	0.0014	0.0043		
2-1-2-1-2	0.0099	0.0167	0.0000	0.0003	0.0019	0.0045
2-2-1-1-2	0.0095	0.0149	0.0002	0.0018		
2-2-1-2-2	0.0002	0.0015	0.0018	0.0047		
1-2-2-1-2	0.0026	0.0076	0.0029	0.0056		
2-2-2-1-1	0.0003	0.0028	0.0017	0.0054		
2-1-1-2-2	0.0036	0.0080	0.0103	0.0138		
1-2-2-2-2	0.0076	0.0105	0.0045	0.0092	0.0009	0.0029
2-1-2-2-2			0.0006	0.0030		
2-1-2-2-1			0.0012	0.0041		
1-2-2-1-2					0.0152	0.0180
2-2-2-2-2	0.0002	0.0014				

Table 3. HLA-Alu insertion frequencies in selected populations used for genetic distance analysis

Population	No Tested	*Loci*					Reference
		AluMICB	***AluTF***	***AluHJ***	***AluHG***	***AluHF***	
British	61	0.271	0.238	0.328	0.308	0.213	Present study
Punjabi Jat Sikh	66	0.208	0.144	0.348	0.182	0.204	Present study
Gujarati Patel	63	0.207	0.191	0.309	0.137	0.167	Present study
Australian	105	0.157	0.107	0.252	0.301	0.203	Kulski and Dunn 2005
Japanese	99	0.118	0.083	0.376	0.27	0.064	Kulski and Dunn 2005
Northeastern Thai	192	0.117	0.086	0.292	0.292	0.018	Dunn et al 2005
Mongolian	41	0.378	0.22	0.293	0.22	0.098	Kulski and Dunn 2005
Southeastern Bantu	50	0.03	0.1	0.07	0.06	0.09	Kulski and Dunn 2005
Khoi	43	NT	0.167	0.186	0.131	0.163	Kulski and Dunn 2005
Sekele San	60	0.1	0.034	0.049	0.033	0.082	Kulski and Dunn 2005
!Kung San	42	0.036	0.238	0.107	0.073	0.06	Kulski and Dunn 2005
Chinese-Yunnan Han	82	0.122	0.055	0.085	0.305	0.018	Yao et al. 2010
Chinese-Jinuo	108	0.106	0.019	0.06	0.273	0.028	Yao et al. 2010
Chinese-Malaysia	50	0.17	0.04	0.3	0.56	0.03	Dunn et al 2007

NT Not tested

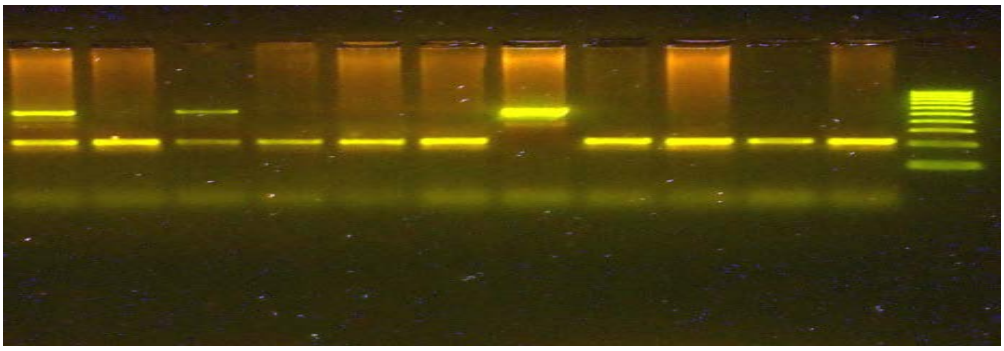
Supporting information

Table 1. Table to show forward and reverse primer sequences for all Alu's

Aluy	**Primer**	**Primer sequence**	**Tm**	**Fragment size -**	**Fragment size +**
HG	For	5'-CAGGACAACCAGTAAAGATGCTGG-3'	60°C	218bp	540bp
	Rev	5'-GCTTCAGTTAACATGCAAGTTTATGCC-3'			
MICB	For	5′-GCCTTCCAATGCCATTCACAG-3′	67°C	503bp	665bp
	Rev	5'-CTCAGCCCTGCTTTCCCATCT-3'			
HJ	For	5'-AAGAAACCCATAACTCACTTG-3'	60°C	173bp	501bp
	Rev	5'-TGTGTCC AGGTTAAACTTCAG-3'			
TF	For	5'-GTGCCTGGTAAAAATTTAAGAGCTGTA-3'	62°C	426bp	710bp
	Rev	5'-TGCACCCGGCCTAAAACCACTGGTT-3'			
HF	For	5'-GCCTCATGGCCTGAATCTGCCAGTGTCCTT-3'	62°C	456bp	605bp
	Rev	5'-GTAACTGACGTGCCCTCTATATAGTATAGTCT-3'			

Supporting Information

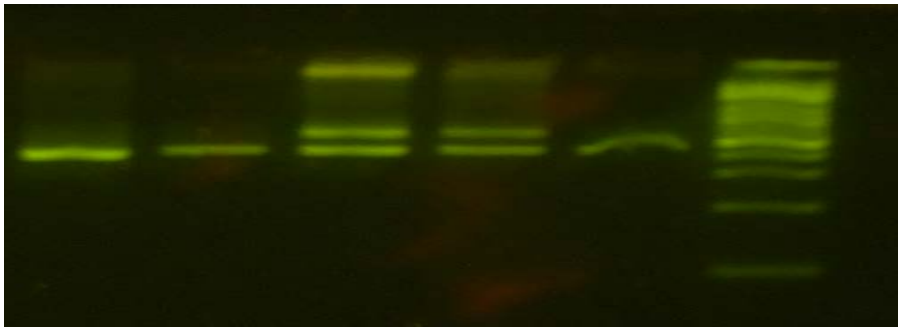
Representative gel pictures (cathode is on the top)



1 2 3 4 5 6 7 8 9 10 11 12

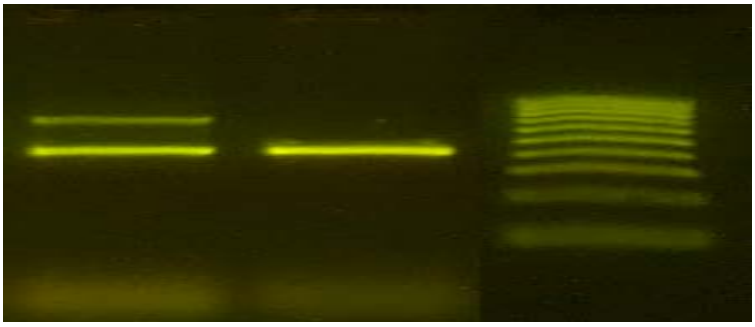
AluyHG gel: (Deletion (D) allele = 218bp, Insertion (I) allele =540bp)

Lane 12: 100 bp ladder, Lanes 1 & 3 heterozygote (ID); Lane 7 homozygote for insertion allele (II) and lanes 2,4-6, and 8-11, homozygote for deletion allele (DD).



1 2 3 4 5 6

Gel photographs for AluTF: (Deletion (D) allele = 426bp, Insertion (I) allele =710bp) Lane 6: 100bp ladder, Lanes 1, 2&5: homozygote for deletion allele (DD) Lanes 3&4 heterozygote (ID) genotypes.



1 2 3

Gel photographs for AluHF: (Deletion (D) allele = 456bp, Insertion (I) allele =605bp) Lane 3: 100bp ladder, Lanes 1: heterozygote (ID) genotypes and Lane homozygote for deletion allele (DD)