A World Wide Survey on Human Specific *Alu* Insertion Polymorphisms

Antti Sajantila
University of Helsinki
Department of Forensic Medicine, P.O. Box 40, 00014,
Helsinki, Finland

ABSTRACT

Alu insertion polymorphisms are widely distributed in the human genome. They are readily analysed and provide an interesting set of markers for human population genetic and forensic studies. The aim of this article is to present preliminary results from our ongoing project on a survey of *Alu* insertion polymorphisms in various human populations.

INTRODUCTION

Repetitive DNA sequences comprise a significant portion of the human genome (Fig 1). *Alu* repeats are classified as short interspersed elements (SINEs) and make up 5-10 % of the human genome (1,2). A typical *Alu* is a heterodimer of two subunits, derived from the 7SL RNA gene. This retroposon invaded the pre-simian and early simian genomes in great copy numbers. Thus, they can serve as milestones in the anthropological and evolutionary studies (Fig. 2).

The insertion of a *Alu* element at a particular locus can be regarded as a unique event. Once inserted, most *Alu* elements do not appear to be subject to loss or rearrangement, therefore being stable genetic markers. *Alu* deletions are rare and even then the deletion leaves behind a footprint (3). The possibility of a deletion, where the two cut points would be exactly at the both ends of the insertion, is negligible.

Human specific Alu insertions

Alu elements that are mostly, but not exclusively restricted to the human genome have been termed human-specific (HS) (4,5) (Fig. 2, Fig 3). The HS *Alu* insertions are further classified into subfamilies Ya5, Ya8 and Yb8 according to base substitutions along their sequence (5). Since the bi-allelic genotyping of *Alu* insertions is feasible using PCR, agarose gel electrophoresis and ethidium bromide staining, they provide an interesting approach for studies of human history. Indeed, the usefulness of HS *Alu* polymorphisms as a tool in human population studies has been proven in previous studies (3, 6, 7).

MATERIALS AND METHODS

We have analysed eight *Alu* insertion polymorphisms (APO, PV92, TPA25, FXIIIB, D1, ACE, A25, and B65) in 57 populations consisting of a total of 2308 unrelated individuals that were sampled from 16 African populations, from 5 Middle Eastern populations, from 12 European populations, from 17 Asian populations and from 3 and 4 populations from Sahlu and New World, respectively. The samples included data from Bengs *et al*. (unpublished) and Stoneking *et al.* (7).

Genomic DNA was prepared from peripheral blood lymphocytes using standard protocols. Eight human specific polymorphic *Alu* loci were amplified using polymerase chain reaction (PCR) and locus specific primers. PCR reactions were carried out in a 25 ml volume containing 20-100 ng of template DNA, 5 nmol of each dNTP, 25 pmol of each primer and 1 unit of *Taq* DNA polymerase (Promega) in 50 mMTris-HCl, pH 8.8, 15 mM $(NH_4)_2SO_2$ 4,1,5 mM $MgCl_2$, 0,1 % Triton X-100, 0,01 % gelatin. 30 cycles of 95 °C for 1 min, 54 °C for 1 min, and 72 °C for 1 min were used in a MJ Research PTL-225 thermal cycler for loci PV92, TPA25, APO, and FXIIIB. 30 cycles of 95 °C for 1 min, 56 °C for 1 min, and 72 °C for 1 min were used in a MJ Research PTL-225 thermal cycler for loci ACE, A25 and B65. Touchdown PCR with annealing temperatures from 61 °C to 58 °C in a total of 30 cycles was used for the D1 locus. PCR products were visualized in UV-light after separation in a 2 % agarose gel and ethidium bromide staining.

Allele frequencies, heterozygosity and G_{ST} values were calculated using GENEPOP (ver. 3) (8) computer package. In addition, calculation of Nei's standard genetic distances (D) between populations, D_A distances between populations and construction of phylogenetic trees were performed using the DISPAN computer program (9). Analysis of molecular variance (AMOVA) was calculated using ARLEQUIN computer package (10).

RESULTS AND DISCUSSION

Hardy-Weinberg equilibrium

All the populations were tested for the Hardy-Weinberg equilibrium (HWE). No violation of the assumption of the HWE could be demonstrated when using the exact test or the global test across all loci in all populations.

Frequency of the Alu insertion and heterozygosity values in various populations

Our results indicate a difference in the frequencies of the *Alu* insertions between various human population groups. For example, at locus PV92 the frequency of the *Alu* insertion varied from 0.08 to 0.91 and at locus TPA25 the frequency varied from 0.19 to 0.56 in the populations studied. Furthermore, at locus APO the frequency of the *Alu* insertion varied from 0.17 to 1.00. It is noteworthy that in Eurasian and Middle Eastern population groups the frequency varied from 0.94 to 1.00, whereas in Sub-Saharan and Saharan population groups the frequency varied from 0.17 to 0.76. Also at locus FXIIIB the frequency of the *Alu* insertion was lower (0.11-0.30) in Sub-Saharan and Saharan Africans compared to that in the Arabs and the Jews of Middle East (0.47-0.56) and to that in the Eurasians (0.37-0.96). Interestingly, at locus D1 the frequency of the *Alu* insertion did not exceed 0.50 in any of the population groups studied.

The heterozygosity values across all loci varied from 0.28 to 0.39 in Eurasian populations, from 0.32 to 0.41 in Arabs and Middle Eastern Jews, and from 0.35 to 0.54 in Sub-Saharan and Saharan African populations.

Genetic distances and construction of phylogenetic trees

The genetic distances between populations showed the largest values between the African and non-African populations. This is also shown in phylogenetic trees constructed by using the neighbor-joining (NJ) method (11) from matrices of either D or DA distances (data not shown). In all the phylogenetic trees constructed, the African populations are closest to the outgroup. As an outgroup we used the genotypes of ancestral state of the *Alu* insertions (non-insertion). Thus, our data supports the findings in phylogenetic trees contructed using other gentic markers, such as mitochondrial DNA, autosomal microsatellites and serological markers.

AMOVA analysis

For the analysis of molecular variance the populations were grouped according to the geographical origin. AMOVA analysis computes a hierarchical analysis of the level of diversity within populations, among populations within the geographical groups and among the geographical groups. The results were then compared to those obtained from literature using other markers. In eight *Alu* markers studied, the majority of the genetic diversity is observed within populations (86 %) whereas the diversity among populations within geographical regions is only in order of 2 %, and the among the geographical groups about 12 %. Our findings are in agreement with our observation of AMOVA analysis of the second hypervariable region at the human mitochondrial DNA (12), and to that of Barbujani *et al.* (13), which studied several STR and RFLP markers (Fig 4).

CONCLUSIONS

Our data shows that *Alu* insertion polymorphisms are very useful markers for population genetic studies. Although showing less variation in a particular locus compared to *e.g* microsatellite loci, at the population level the level of genetic diversity detected in these markers are in accordance with other markers utilised in forensic science. Thus, they would be useful marker systems as models for populations genetic analyses utilised in forensic science.

ACKNOWLEDGEMENTS

I am grateful for Auli Bengs, Kirsti Höök, Marjo Leppälä, Nira Gallili, Elizabeth Watson, Andrei Kozlov, Mark Batzer, Steven Sherry and Mark Stoneking for continuous help, providing samples and unpublished data for the project.

REFERENCES

1. Arcot S.S., Adamson A.W., Lamerdin J.E., Kanagy B., Deininger P.L., Carrano A.V., and Batzer M. (1996). *Alu* fossil relics - distribution and insertion polymorphism. *Genome Research* **6**:1084-1092.

2. Maraia R.J. (ed). The Impact of Short Interspersed Elements (SINEs) on the Host Genome. Springer-Verlag, Heidelberg, Germany, 1995.

3. Stoneking M. (1997) The human genome project and molecular anthropology. *Genome Research* **7**:87-91.

4. Batzer M.A., Kilroy G.E., Richard P.E., Shaikh T.H., Desselle T.D., Hoppens C.L., and Deininger P.L. (1990). Structure and variability of recently inserted *Alu* family members. *Nucleic Acids Res*. **18**: 6793-6798.

5. Batzer M.A. and Deininger P.L. (1991) A human-specific subfamily of *Alu* sequences. *Genomics* **9**:481-487.

6. Batzer M.A., Stoneking M., Alegria-Hartman M., Bazan H., Kass D.H., Shaikh T.H., Novick G.E., Ionnou P.A., Scheer W.D., Herrera R.J., and Deininger P.L. (1994) African origin of human-specific *Alu* insertions. *Proc Natl Acad Sci* **91**:12288-12292.

7. Stoneking M., Fontius J.J., Clifford S.L., Soodyall H., Arcot S.S., Saha N., Jenkins T., Tahir M.A., Deiniger P.L., Batzer M.A. (1997) *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Research* **7**: 1061-1071.

8. Raymond M. and Rousset F. (1995). GENEPOP (version1.2): population genetics software for exact tests and ecumenicism. *J. Heredity* **86**:248-249.

9. Ota T. (1993). DISPAN: Genetic distance and phylogenetic analysis. Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, USA.

10. Schneider S., Kueffer J.-M., Roessli D., Excoffier L. (1997) ARLE-QUIN ver1.1 - A software for population genetic data analysis.

11. Comas D., Reynolds R., Sajantila A.. Analysis of mtDNA HVRII in several human populations using an immobilised SS probe hybridisation assay. *Eur J Hum Genet*, in press

12. Barbujani G., Magani A., Minch E., Cavalli-Sforza L.L.. An apportionment of human DNA diversity. *Proc Natl Acad Sci USA* **94**: 4516-4519.

13. Saitou N,. Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406-425.

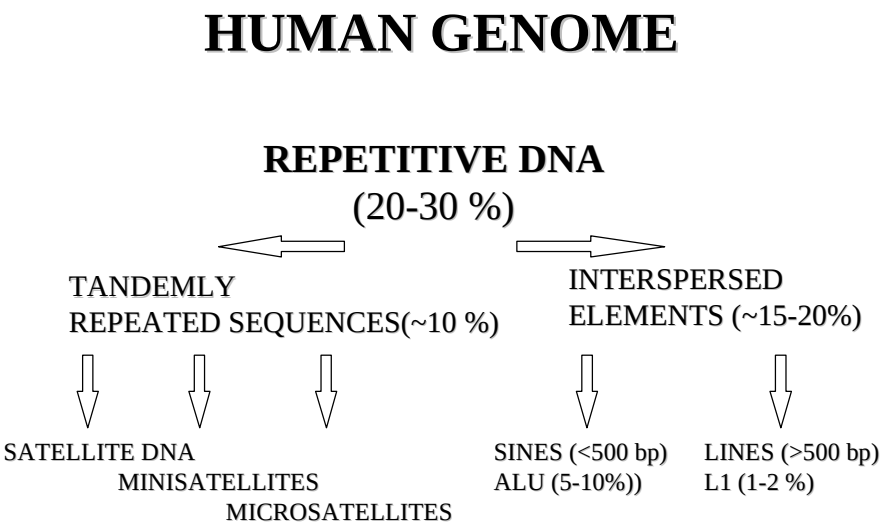

Figure 1. Schematic organisation of repetitive DNA in the human genome. *Alu* insertion polymorphisms constitute approximately 5-10 % of the human genome.

ALU INSERTIONS AS EVOLUTIONARY MILESTONES

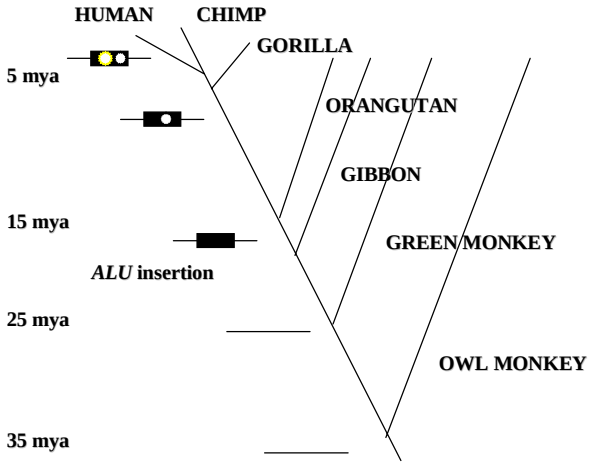

Figure 2. *Alu* insertions are derived from the 7SL RNA gene. The species specific *Alu* insertions have gained characteristic base substitutions during evolutionary time. Thus, they can be classified to *e.g.* primate specific, anthropoid specific and human specific *Alu* insertions. Due to this characteristic, they can serve as unique evolutionary milestones.

CLASSIFICATION OF *ALU* INSERTIONS

- consensus sequence ~ 300 bp

MOLECULAR MILESTONES

- primate specific (PS)
- anthropoid specific (AS)
- catarrhine specific (CS)
- human specific (HS-1, HS-2)
- Sb 2

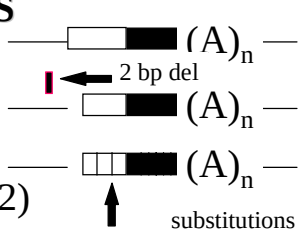

Fig 3. Classification of *Alu* insertions. A typical *Alu* is a heterodimer of two subunits, derived from the 7SL RNA gene. *Alu* elements that are mostly, but not exclusively restricted to the human genome have been termed human-specific (HS) (4,5) (Fig. 2, Fig 3). The HS *Alus* are further classified into subfamilies Ya5, Ya8 and Yb8 according to base substitutions along their sequence.

AMOVA

MARKER	WITHIN POPULATIONS	AMONG POPULATIONS WITHIN GROUPS	AMONG GROUPS
8 ALUs	**86.2 %**	**1.9 %**	**11.9 %**
mtDNA(HVII) (Comas *et al*. Eur J Hum Genet, in press)	**93.1 %**	**3.8 %**	**3.1 %**
30 STRs	**84.5 %**	**5.5 %**	**10.0%**
16 RFLPs	**83.6 %**	**8.4 %**	**8.0 %**

Table 1. Comparison of AMOVA analysis of *Alu* insertions compared to that of mtDNA, STRs and RFLPs.