 IJHG 1(4): 249-254 (2001)

Alu Insertions and Ethnic Composition in a Brazilian Population Sample

Celso Teixeira Mendes-Junior and Aguinaldo Luiz Simões

Department of Genetics, School of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP, Brazil

KEY WORDS *Alu* insertion; PSAs; polymorphism; admixture; paternity.

ABSTRACT *Alu* insertions can be defined as a family of elements approximately 300 base pairs long, which have more than 500,000 copies distributed throughout the human genome. Once inserted in certain loci, these sequences apparently do not suffer any loss or rearrangements. Some of these sequences were inserted recently, such that the presence or absence of the insertion is polymorphic. Three of these markers (TPA 25, PV 92 and APO) were studied in a sample of 149 trios (mother, child, and supposed father) from an urban population from southwestern Brazil (Ribeirão Preto, SP). The goal was to characterize the ethnic composition of this population and evaluate the usefulness of *Alu* insertions in the investigation of paternity. The analysis was done with PCR followed by polyacrylamide gel electrophoresis stained with silver nitrate. The frequencies obtained, intermediate between African and European, exhibit Hardy-Weinberg equilibrium, and they reflect the composition of an urban population to which European and African ethnic components made a greater contribution than Amerindians. Our results corroborate previous reports concerning the difficulty of amplification of these markers, contraindicating their use for forensic purposes.

INTRODUCTION

Alu insertions - thus called because they were initially described as a fraction of repetitive DNA that exhibited an *Alu* enzyme restriction site (Houck et al. 1979) - are sequences approximately 300 base pairs long found exclusively in primates. They occur every 3-6 kbp (Mighell et al. 1997), amounting to more than 500,000 copies found throughout the human genome (Batzer et al. 1994). New copies of these elements insert themselves in the genome at a rate of approximately 100 to 200 per million years (Batzer et al. 1994) via reverse transcription of an intermediate RNA in a process known as retroinsertion; all of the insertions of a locus are identical by descent, because it is highly improbable that the insertion phenomenon occurs twice at the same locus (Batzer and Deininger 1991). There is no mechanism for the precise removal of an insertion; when this occurs, a "signal" indicating the original insertion event is left behind (Batzer et al. 1994).

Some of these insertions, by having retroposed so recently in human evolutionary history, are still polymorphic for presence/absence at a specific location (Stoneking et al. 1997). However, the ancestral state of the polymorphism is the absence of the insertion, a fact supported by the existence of human-specific *Alu* insertions; one may conclude that the direction of the mutational changes is the gain of the *Alu* element at a given locus. The characteristics of this polymorphism, the possibility of determining its ancestral state and the direction of the mutational changes, not possible for other markers, makes it unique with respect to its use in the analysis of relationships between populations; references have been made with respect to their potential applications, such as in paternity investigation cases (Batzer et al. 1994; Novick et al. 1995; Thomas and Herrera 1998). Population samples from distinct geographic origins have shown great differences in the allele frequency distributions of these polymorphisms (Batzer and Deininger 1991; Stoneking et al. 1997); recently, some of these elements have been identified as "population-specific alleles" (PSAs) since they present large frequency differentials between two well defined populations (Shriver et al. 1997), suggesting that these insertions could be excellent racial markers and, thus, of great use in estimating ethnic composition of hybrid populations like Brazil's, which is comprised of Blacks, Europeans, and Amerindians.

Complete Address (correspondence): Aguinaldo Luiz Simões, Department of Genetics, School of Medicine of Ribeirão Preto, University of São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, SP Brazil - CEP: 14049–900 *Phone:* +55 16 6023157, *Fax:* +55 16 6330069, *E-mail:* alsimoes@fmrp.usp.br

In the first 350 years after the start of Portuguese colonization of Brazil, to the end of the slave trade in 1850, 2 million indigenous people mixed with approximately 3.6 million blacks and 1.9 million Portuguese; in the 1890 census, the Brazilian population was estimated at close to 14.7 million people (2 million blacks, 440 thousand Indians, and the rest were approximately equally divided between whites and mixed). In the remaining years, up to the present, the entrance of 5 million Europeans and 260 thousand has been recorded (Callegari-Jacques and Salzano 1999). In the current population of approximately 170 million, the proportion of whites predominates strongly in the South of the country, and it is nearly always greater than the other two ethnic groups in the rest of the country except for the black enclave in Bahia; in the Amazon region, the second greatest component is Amerindian, while in the Northeast, this position is occupied by an African contingent.

We determined the allele frequencies of three *Alu* insertion-type markers (TPA 25, PV 92 and APO) in an urban population from southwestern Brazil. Two of them have shown large frequency differentials: the locus PV 92 presents differences between Amerindian/African and Amerindian/European populations (Shriver et al. 1997), while the locus APO presents between African/European (Parra et al. 1998). The TPA 25 locus is the only one that shows homogeneity in frequencies distribution worldwide. The genotyping was done using PCR and polyacrylamide electrophoresis with silver nitrate staining, in place of the traditional method of agarose gel electrophoresis containing ethidium bromide. The sample, originating from trios undergoing paternity investigation, allowed for the testing of the use of these markers for this specific purpose.

MATERIALS AND METHODS

The population sample was composed of individuals involved in disputed paternity cases investigated by the Department of Genetics of the University of São Paulo Hospital at Ribeirão Preto, São Paulo between July 1996 and August 1997. The subjects, involving 149 trios, were mainly white, and they were from the surrounding area. All of them were subjected to a routine test in which several STRs were studied to calculate a probability of paternity equal to or greater than 99.99% or exclusion in at least two loci (alleged paternity was proved to be false in 32 of the 149 cases).

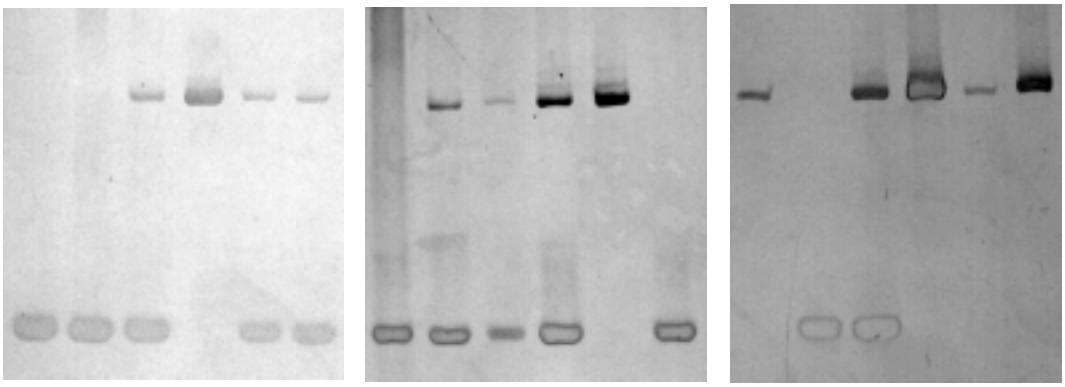
The TPA 25 (Degen et al. 1986), PV 92 (Batzer et al. 1994) and APO (Batzer et al. 1991) insertions are present exclusively in the human genome, and they are located on chromosomes 8, 16, and 11, respectively. DNA extracted from whole blood (Higuchi, 1989) was amplified by PCR according to Batzer et al. (1996), and amplification products were separated by non-denaturing 6% polyacrylamide gel electrophoresis (PAGE) followed by silver staining (approximately 30 samples were analysed as well with agarose gel with the same results). Some of the tests were repeated with modification (addition of dimethyl sulfoxide - DMSO) according to Odawara et al. (1997).

The expected probability of exclusion was calculated with the Weir formula (Weir 1996), and the ethnic composition was determined using the method of maximum likelihood estimation (Krieger et al. 1965). Mathematical analyses were performed using the GENIOC program (Cabello and Krieger 1997).

RESULTS AND DISCUSSION

This is the first estimate of *Alu* insertion frequencies in South American urban populations. Under the conditions employed, amplification of the sequences that contain the insertions produced fragments with 400bp (Fig. 1); in their absence, the fragments produced were 110bp (Novick et al. 1995).

Due to the earlier described difficulty of amplification of the *i* allele (presence of the insertion) in relation to the *a* allele (absence of the insertion) (Odawara et al. 1997; Perna et al. 1992), the non-amplified samples (*na*) or those determined to be homozygous for insertion absence (*aa*) were reanalysed at least two times; in the case of TPA 25, we used DMSO in the reaction mixture with results similar to those repetitions without this reagent. One reading of *ai* (heterozygous) was considered sufficient for definition of the phenotypes, even if the same



A **B** **C**

Fig. 1. PAGE, 6% non-denaturing, stained with silver nitrate. A. Locus TPA 25; B. Locus PV92; C. Locus APO. In the LANE A4, B5, C1, C4, C5 and C6, *ii* homozygous; in the LANE A3, A5, A6, B2, B3, B4 and C3, heterozygous and in the LANE A1, A2, B1, B6 and C2, homozygous *aa*..

Table 1: Comparison of the results from the first analysis with those obtained after repeat analyses in the initial non-amplified samples (*na*) or classified as homozygous for the absence of the insertion (*aa*)

Loci	*First diagonosis*	*Final diagnosis*			
		aa	*ai*	*ii*	*na*
TPA 25					
	114 aa	80	34	-	-
	16 na	2	6	4	4
PV 92					
	218 aa	176	42	-	-
	3 na	-	1	1	1
APO					
	17 aa	8	9	-	-
	8 na	-	2	6	-

sample had had several *aa* readings.

The reanalysis significantly changed our results (Table 1), correcting the excess of *aa* homozygotes observed initially and the underestimate of the insertion frequencies (42%, 14% and 85% before and 49%, 21% and 86% after for the loci TPA 25, PV 92 and APO, respectively); in the end, the estimated genotype frequencies were in Hardy-Weinberg equilibrium (Table 2).

The results show that PAGE is a technique applicable to the study of these markers. Our results also confirm the difficulty of amplifying the allele containing the insertion (Odawara et al. 1997; Perna et al. 1992) in relation to the other. However results do not confirm the hypothesis

Table 2: Gene and genotype frequencies, Hardy-Weinberg equilibrium, and power of exclusion (observed and expected) of the loci *Alu* TPA 25, PV 92 and APO in a sample from a southeastern Brazilian population.

	TPA 25	*PV 92*	*APO*
aa	82	176	8
ai	135	102	65
ii	75	11	219
na	4	1	0
N	296	290	292
Alu insertion	0.488	0.215	0.861
χ^2	1.634	0.645	1.362
p	0.2012	0.4219	0.2432
Power of Exclusion (PE)	0.1874	0.1403	0.1054
Exclusions Observed	0.2500(6/24)	0.1034(3/29)	0(0/29)

(Odawara et al., 1997) that the majority of the non-amplified samples would be, in fact, *ii* (homozygous for presence of the insertion).

Of the 149 trios analysed, there was exclusion of alleged paternity in 32 cases, via routine tests. The number of exclusions confirmed by the 3 *Alu loci* (Table 2) did not significantly differ from the estimate of expected power of exclusion ($p = 0.4319$, $p = 0.5686$ and $p = 0.0645$, for the *loci* TPA 25, PV 92 and APO, respectively). In 3 cases (Table 3), the STRs indicated possible paternity with probability higher than 99.99%, but TPA 25 revealed a maternal exclusion (*aa* mother, *ii* child and *ii* father) and two paternal exclusions (*ai*

Table 3: Comparison between the results from the investigation of paternity using routine tests (STRs) and those obtained with the three markers of *Alu* insertions in 149 trios. *na* = non-amplified.

STR	*Alu*	*TPA 25*	*PV 92*	*APO*
Possible	Possible	106	114	108
Exclusion	Possible	18	26	29
Possible	Exclusion	3	-	-
Exclusion	Exclusion	6	3	-
Possible	na	8	3	9
Exclusion	na	8	3	3

child and *aa* mother and father). Such readings did not change even after 4 or more repetitions. The difficulty of correct amplification may be an explanation for these three cases in which there was maternal or paternal exclusion; other hypotheses, such as the presence of a null allele or an elevated mutation rate are not very likely. This problem of phenotype classification makes it difficult to accept the use of these systems for forensic purposes.

The genetic frequencies in our sample exhibit values that are intermediate between those described for Europeans and Africans (Fig. 2), reflecting the contribution of these two ethnic groups to Brazil's population. To estimate ethnic mixture, the weighted means from the frequencies reported in the literature for some populations of the African and European continents were considered as ancestral gene frequencies (Batzer et al. 1996; Batzer et al. 1991; Batzer et al. 1994; Comas et al. 2000; Parra et al. 2001; Parra et al. 1998; Stoneking et al. 1997; Tishkoff et al. 1996; van der Bom et al. 1997; Watkins et al. 2001); for the Brazilian Amerindians, in addition to data from the literature (Novick et al. 1998; Tishkoff et al. 1996), unpublished data (Oliveira, 1999 - personal communication) obtained from our laboratory were used (Table 4). The data presented here indicate that for the southeastern Brazilian population, the relative contribution of the indigenous ancestrals was null or insignificant, while those related to the European and African ancestrals were estimated at 78% and 22%, respectively. The same was observed when considering only the two PSAs (79,5% and 20,5%, respectively).

These results are somewhat different to that estimated in another sample from Ribeirão Preto

Table 4: Ancestral gene frequencies used in this study for estimating ethnic mixture in a Brazilian population sample.

Alu insertion	*Europeons*	*Africans*	*Amerindians*
TPA 25	0.578	0.252	0.599
PV 92	0.225	0.255	0.919
APO	0.952	0.514	0.994

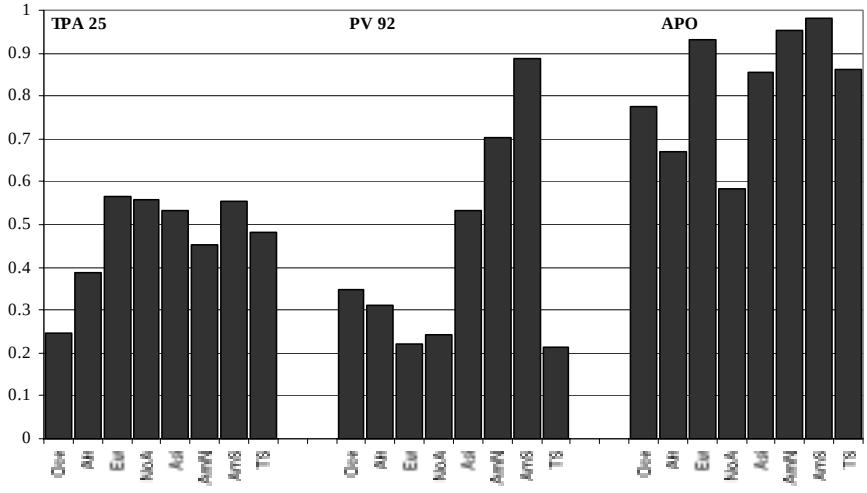


Fig 2. World distribution of the weighted mean frequencies of the *Alu* insertions TPA 25, PV 92 and APO (Batzer et al., 1996; Batzer et al., 1991; Batzer et al., 1994; Comas et al., 2000; Melton et al., 1998; Novick et al., 1998; Novick et al., 1995; Parra et al., 2001; Parra et al., 1998; Perna et al., 1992; Stepanov et al., 1999; Stoneking et al., 1997; Tishkoff et al., 1996; van der Bom et al., 1997; Watkins et al., 2001). Oce: Oceania; Afr: Africa; Eur: Europe; NoA: North America; Asi: Asia; AmN: Amerindians-North America; AmS: Amerindians-South America; TS: This Study.

and it's surrounding area (Leboute, 2001 – personal communication) using three autosomal microsatellites (68,25% European, 14,96% African and 16,79% Amerindian). The PSAs estimates are in agreement with historical data, which indicates a minimum of Amerindian contribution to the formation of the population here analysed, suggesting that these markers have high informativeness to admixture studies, although additional comparative approaches are necessary.

ACKNOWLEDGEMENTS

We are grateful to Mrs. Maria do Carmo Tomitão Canas and Ana Lucia Pimentel for technical assistance in laboratory analyses. This research was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

REFERENCES

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, Gudi VA, Mena JC, Foltz DW, Herrera RJ, Deininger PL 1991. Amplification dynamics of human-specific (HS) Alu family members. *Nucleic Acids Res*, **19:** 3619-3623.

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL 1994. African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci USA*, **91:** 12288-12292.

Cabello PH, Krieger H 1997. *Genioc: sistema para análise de dados de genética*. Rio de Janeiro: Fundação Osvaldo Cruz.

Callegari-Jacques SM, Salzano FM 1999. Brazilian Indian/non-Indian interactions and their effects. *Ciência e Cultura: Journal of the Brazilian Association for the Advancement of Science*, **51:** 166-174.

Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A 2000. Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet*, **107:** 312-319.

Degen SJ, Rajput B, Reich E 1986. The human tissue plasminogen activator gene. *J Biol Chem*, **261:** 6972-6985.

Higuchi R 1989. Simple and rapid preparation of samples for PCR. In: Henry A Erlich (Ed.): *PCR Technology: Principles an Applications for DNA Amplification*. New York: Stockton Press pp. 31-38.

Houck CM, Rinehart FP, and Schmid CW 1979. A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol*, **132:** 289-306.

Krieger H, Morton NE, Mi MP, Azevedo E, Freire-Maia A, Yasuda N 1965. Racial admixture in north-eastern Brazil. *Ann Hum Genet*, **29:** 113-125.

Melton T, Clifford S, Martinson J, Batzer M, Stoneking M 1998. Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *Am J Hum Genet*, **63:** 1807-1823.

Mighell AJ, Markham AF, Robinson PA 1997. Alu sequences. *FEBS Lett*, **417:** 1-5.

Novick GE, Novick CC, Yunis J, Yunis E, Antunez de Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ 1998. Polymorphic Alu insertions and the Asian origin of Native American populations. *Hum Biol*, **70:** 23-39.

Novick GE, Novick CC, Yunis J, Yunis E, Martinez K, Duncan GG, Troup GM, Deininger PL, Stoneking M, Batzer MA, Herrera RJ 1995. Polymorphic human specific Alu insertions as markers for human identification. *Electrophoresis*, **16:** 1596-1601.

Odawara M, Matsunuma A, Yamashita K 1997. Mistyping frequency of the angiotensin-converting enzyme gene polymorphism and an improved method for its avoidance. *Hum Genet*, **100:** 163-166.

Parra EJ, Kittles RA, Argyropoulos G, Pfaff CL, Hiester K, Bonilla C, Sylvester N, Parrish-Gause D, Garvey WT, Jin L, McKeigue PM, Kamboh MI, Ferrell RE, Pollitzer WS, Shriver MD 2001. Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am J Phys Anthropol*, **114:** 18-29.

Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD 1998. Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet*, **63:** 1839-1851.

Perna NT, Batzer MA, Deininger PL, Stoneking M 1992. Alu insertion polymorphism: a new type of marker for human population studies. *Hum Biol*, **64:** 641-648.

Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, Ferrell RE 1997. Ethnic-affiliation estimation by use of population-specific DNA markers. *Am J Hum Genet*, **60:** 957-964.

Stepanov VA, Puzyrev VP, Spiridonova MG, Khitrinskaia, II 1999. Analysis of the Alu insertion polymorphism in urban and rural populations of Siberia. *Genetika*, **35:** 1138-1143.

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.

Thomas E, Herrera RJ 1998. Multiplex polymerase chain reaction of Alu polymorphic insertions. *Electro-*

phoresis, **19:** 2373-2379.

Tishkoff SA, Ruano G, Kidd JR, Kidd KK 1996. Distribution and frequency of a polymorphic Alu insertion at the plasminogen activator locus in humans. *Hum Genet*, **97:** 759-764.

van der Bom JG, de Knijff P, Haverkate F, Bots ML, Meijer P, de Jong PT, Hofman A, Kluft C, Grobbee DE 1997. Tissue plasminogen activator and risk of myocardial infarction. The Rotterdam Study. *Circulation*, **95:** 2623-2627.

Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen SV, Batzer MA, Harpending HC, Rogers AR, Jorde LB 2001. Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *Am J Hum Genet*, **68:** 738-752.

Weir BS 1996. *Genetic Data Analysis II*. Sunderland: Sinauer Associates Inc.