

Alan D. Lovell · Vania Yotova · Fengxia Xiao
Mark A. Batzer · Damian Labuda

Polymorphisms within a polymorphism: SNPs in and around a polymorphic *Alu* insertion in intron 44 of the human dystrophin gene

Received: 5 December 2003 / Accepted: 10 February 2004 / Published online: 8 April 2004
© The Japan Society of Human Genetics and Springer-Verlag 2004

Abstract A polymorphic Yb-type *Alu* insertion on Xp21.3 shows a genotypic gradient across worldwide populations. We used single strand conformational polymorphism (SSCP), denaturing high-pressure liquid chromatography (DHPLC), and sequencing to characterize the level of polymorphism within this region. Two novel polymorphic sites were found within the *Alu* insertion itself, and a further seven novel polymorphic sites in the 2-kb flanking region. Our results showed that while DHPLC was more sensitive than SSCP, the limitations of DHPLC included the lack of ability to distinguish between multiple alleles or safely identify mutations on a polymorphic background. We believe that this is the first report of polymorphic single nucleotide polymorphisms (SNPs) within a polymorphic *Alu* distribution and that together they promise to provide a useful marker for human population and evolutionary genetics.

Keywords *Alu* insertion · Polymorphism · Demography · Evolution · DHPLC · SSCP · SNP · X-chromosome

Introduction

Polymorphic *Alu* markers often show patterns of diversity whereby their highest frequencies are found

A. D. Lovell · V. Yotova · F. Xiao · D. Labuda (✉)
Centre de Recherche, Hôpital Sainte-Justine,
3175 Côte-Ste-Catherine, Montréal, QC, Canada, H3T 1C5
E-mail: damian.labuda@umontreal.ca
Tel.: +1-514-345-4931/3586
Fax: +1-514-345-4731

M. A. Batzer
Department of Biological Sciences, Biological Computation and
Visualization Center, Louisiana State University,
202 Life Sciences Bldg, Baton Rouge, LA 70803, USA

D. Labuda
Département de Pédiatrie, Hôpital Sainte-Justine,
3175 Côte-Ste-Catherine, Montréal, QC, Canada, H3T 1C5

outside Africa (Watkins et al. 2003). This observation raises questions pertaining to the origin of *Alu* diversity in worldwide populations, though it could simply be due in part to ascertainment bias or lineage sorting of an ancestral polymorphism (Watkins et al. 2003). An *Alu* insertion, classified as type Sb and subsequently Yb (Batzer et al. 1996), and hence of intermediate age (Zietkiewicz et al. 1994), is found 233,908 bp downstream from the informative *dys44* haplotype (Zietkiewicz et al. 2003) in intron 44 on Xp21.3 (Blonden et al. 1994). In order to characterize the diversity of this genomic region, we searched for polymorphic sites both within and flanking the *Alu* insertion using SSCP (single strand conformational polymorphism, Orita et al. 1989) and DHPLC (denaturing high-pressure liquid chromatography), two mutation detection techniques often used to reveal mutations of interest in a candidate gene or region of DNA (Dobson-Stone et al. 2000; Gross et al. 1999). In addition to searching for polymorphic sites with DHPLC and SSCP, we also sequenced in full all *Alu* insertions.

Materials and methods

Eight samples from each of five regions of the world (Sub-Saharan Africa, East Asia, the Americas, the Middle East, and Europe) were taken to avoid ascertainment bias. Of the 40 samples, 12 were males, giving a total of 68 chromosomes. DHPLC analysis requires that hemizygous male DNA is mixed with another sample during PCR to ensure a mixture of heteroduplexes and homoduplexes. Hence, all male DNA was mixed with control DNA (a European male that was not included in the original 40 samples) in equal amounts. Seven pairs of primers were designed using the online Primer3 software (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) to amplify a 2,328-bp region around the 305 bp *Alu* insertion (NCBI accession number Z22650); PCR products ranged in size from 321 to 349 bp to allow satisfactory analysis by both SSCP and DHPLC

Table 1 PCR primer pairs and product length, polymorphic sites, type of polymorphism, minor allele frequency, and the geographic distribution of heterozygotes (in counts) for each polymorphic site. SSCP single strand conformational polymorphism, DHPLC denaturing high-pressure liquid chromatography

Fragment	Forward primer	Reverse primer	Length (bp)	Start ^a (bp)	End ^a (bp)	Polymorphic site ^b	Mutation	Minor allele frequency (q_o)	Across continents ^b					SSCP ^c	DHPLC ^c	
									Africa (3)	Asia (8)	Americas (7)	M-East (5)	Europe (5)			
Pr_1	GCACCTTATGAAG- TATATGAAAGC	AAACACAAACAA- ATAAGGTAGC	323	-780	-457 -648		t → g t → c + t	0.330 0.015 0.015	3 11 1	3	3	3		✓ ✓ ×	✓ ×	
Pr_2	AAATGTGCAATAC- ATAGCTAAC	GGGAAATAAAGG- GAAAAATGC	334	-495	-161 -484		t → a + a	0.108 0.064	1 1	2	3		✓ ✓ ✓	✓ ✓ ✓	✓ ✓ ✓	
Pr_3 _{Alu}	GAATGTCAGAAC- AAATTAGAGC	CCATATATAGTGG- GTAAAAGATACC	341	-201	140 -151	Alu 98	g → a t → c	0.250 0.087	2 2	1 2	4	2	1	✓ ✓	✓ ×	✓ ✓
Pr_4	TGATTTTGGTATC- TTTTACCC	GTTTTCTCTATCA- CTTTTAGTGG	349	108	457 185	Alu 213	t → c t → c	0.043 0.015	1					×	✓	✓
Pr_5	CGTATGCTCTTCA- TTCTAAAATCC	AGCATGACATCCA- GTTGC	331	410	741 602		g → t	0.047	1	2				×	✓	✓
Pr_6	GTCCATTATTGCC- AGAGG	TTCTGAATGATGG- AACAAAGG	329	681	1010 -		-	-						-	-	-
Pr_7	TTTGTGCTCCTT- GTTC	GGAGAATGTTTCAT- CAGACTGG	321	981	1302 -		-	-						-	-	-

^aPrimer sequence start and end points and the location of polymorphic sites are given as the number of base pairs up- or downstream from the site of the *Alu* insertion. Polymorphic sites found in the *Alu* insertion are notated as the number of base pairs from the start of the *Alu* insertion

^bFrequency data across continents is shown as counts of female heterozygotes. Sample sizes (number of females) are shown in parentheses

^cPolymorphic sites were either detected (✓) or missed (×) by SSCP and DHPLC

(Table 1). PCR reactions contained: 1x buffer, 250 μM dNTPs, 2.2 mM/Mg²⁺, 0.6 μM primers, and 1.25 U of Taq polymerase (GibcoBRL). Thermocycler conditions consisted of 35 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 30 s, and elongation at 72°C for 30 s, for all primer pairs.

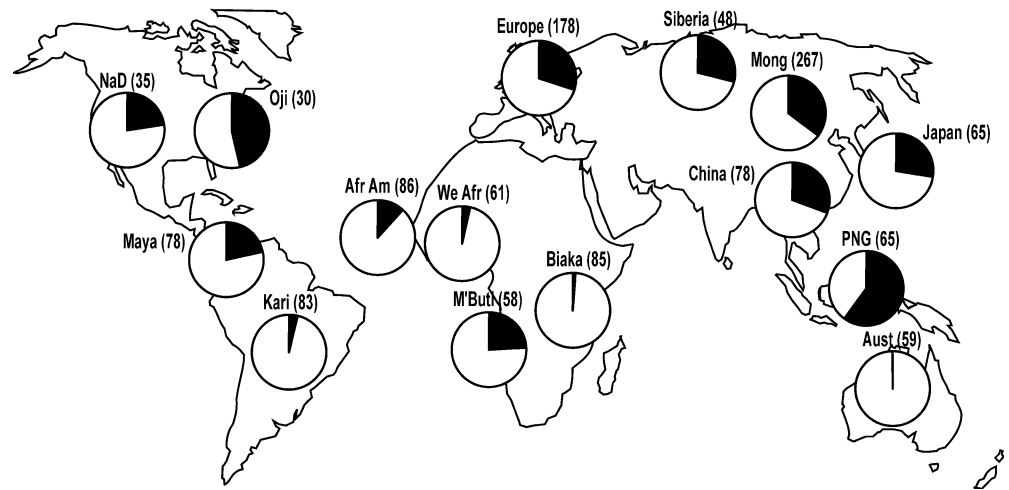
SSCP (Orita et al. 1989; Zietkiewicz et al. 1997) and DHPLC were run in parallel with the two operators working blindly. SSCP was used following previously described procedures (Zietkiewicz et al. 1997) with a 6% polyacrylamide gel containing 10% glycerol (acrylamide-bisacrylamide ratio of 29:1) in TBE and run at room temperature for 16–21 h. For DHPLC (using a Trans-Genomic WAVE 3,500 DHPLC), a gradient of temperatures was run around the (manufacturer-provided) software-predicted optimal temperatures for three or four of the 40 PCR products for each fragment, and the two temperatures that showed the greatest split between heteroduplexes and homoduplexes (when present) were used for each sample-fragment set (to allow the resolution of mutations in different parts of the sequence). All differing SSCP and/or DHPLC patterns, indicating the presence of polymorphic sites, were identified for each run, and at least two individuals were sequenced from each group of individuals displaying a particular pattern; all *Alu* insertions (a total of 19 chromosomes) were sequenced manually as a matter of course. For each identified polymorphic site, the frequency of the minor allele among females samples (q_f) was estimated assuming Hardy–Weinberg equilibrium by $q_f = (1 - \sqrt{1 - 2H})/2$, where H is the frequency of heterozygotes from the DHPLC data for the 28 female samples at each polymorphic site, while the frequency of the minor alleles among the hemizygous males (q_m) was from allele counting. Overall minor allele frequency (q_o) was therefore estimated as $q_o = (q_f \times 0.56)(q_m \times 0.12)/68$, where 56, 12, and 68 are the numbers of female, male, and total number of chromosomes, respectively.

Results and discussion

The global distribution of the polymorphic *Alu* insertion among worldwide samples from Zietkiewicz et al. (2003) and Australian Aboriginal samples from Perna et al. (1992) is shown in Fig. 1. It is of interest to note that the *Alu* insertion reaches its highest frequency in Papua New Guinea but was not found at all in 59 unrelated Australian Aborigine chromosomes. Within Africa, the *Alu* insertion showed a high frequency in the M'Buti Pygmies but is nearly absent in the Biaka, while the presence of *Alu* in African Americans could be due to admixture with European or indigenous populations or inherited from their ancestral West African populations. Within Eurasia, the frequency of *Alu* is fairly consistent, while in the Americas, its frequency varies widely.

A total of seven novel polymorphic sites were found in the 2,328-bp flanking region of the site of the *Alu* insertion; one triallelic single nucleotide polymorphism

Fig. 1 Worldwide *Alu* frequencies (sample sizes, as number of chromosomes, shown in parentheses). All samples are from Zietkiewicz et al. (2003), except the Australian Aborigines, which are from Perna et al. (1992). *Afr Am* African Americans, *Aust* Australian Aborigines (Perna et al.), *Kari* Karitiana, *Mong* Mongolian (all grouped together from Zietkiewicz et al.), *NaD* NaDene, *Oji* Ojibwa, *PNG* Papua New Guinea (highland and coastal), *We Afr* West Africans (all grouped together from Zietkiewicz et al.)



(SNP), four biallelic SNPs, and two single base insertions. Both SSCP and DHPLC identified the insertions and the triallelic SNP, though neither method convincingly identified it as triallelic (it was scored to be biallelic). Of the four biallelic SNPs, all were identified by DHPLC but only one by SSCP. Within the *Alu* insertion itself, one SNP was identified by SSCP but was missed by DHPLC; a second site was missed by both techniques and only uncovered by sequencing. The poor level of detection within the *Alu* insertion was likely caused by the increased fragment length due to the presence of the 305-bp insert (a total of 646 bp, with the flanking fragment Pr_3_ *Alu*).

DHPLC appears to be limited by its ability to score a mutation on a polymorphic background (Dobson-Stone et al. 2000), though Gross et al. (1999) argued that they were able to call such sites confidently. Fragments Pr_1 and 2, which both have one SNP and one single base pair insertion polymorphism, displayed very strong DHPLC patterns for individuals with an insertion (Fig. 2a,b) such that they would overshadow the “shoulder” pattern that was present for individuals heterozygous at the SNP allele thus making the identification of the state of the SNP difficult. Furthermore, it was also not possible to detect or genotype SNPs when more than two alleles were present at the same site. Position -648 was found to have three possible alleles (g, t, or c), and while the heterozygote was identified by the presence of a shoulder on the DHPLC profile, there was no unequivocal and repeatable difference between a “t/g” heterozygote and a “c/g” heterozygote (Fig. 2a).

Assuming Hardy-Weinberg equilibrium, we were able to use the combined female and male data to estimate the overall minor allele frequency (q_o) for each polymorphic site. The highest values of q_o were found for the t → g SNP at position -648 (allele frequency 0.33), and the g → a polymorphism at position -151 (0.25); these two polymorphisms were also found worldwide (i.e., in each of the five geographic regions. Although not shown in Table 1, the t → g SNP at position -648 was found in two European males). Three

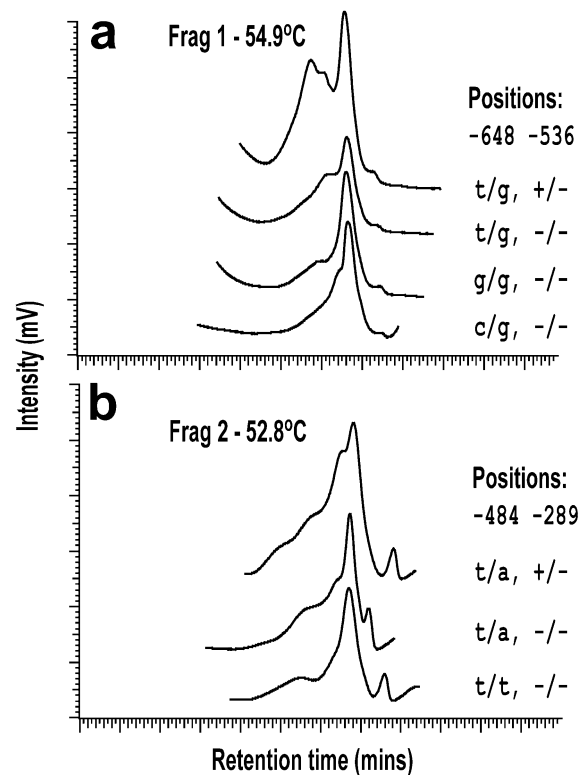


Fig. 2a,b DHPLC profiles highlighting the difficulty of using this approach for detecting mutants on an already polymorphic background. **a** fragment Pr_1 and **b** fragment Pr_2. For each profile, the known genotypes (after sequencing) are shown as bases for the single nucleotide polymorphism (SNP) and presence (+) or absence (-) of the single base pair insertion

flanking region polymorphic sites appeared only once within the sample: the two t → c SNPs at positions -648 and +185 and the +t insertion at position -536. It is notable that all polymorphisms within the flanking region were found at least once within Africa. The two novel polymorphic sites found in the *Alu* insertion itself were both t → c SNPs, at positions 98 and 213, respectively, as measured from the beginning of the *Alu*

insertion; the SNP at position 98 was found twice, both times in Chinese individuals, while the SNP at position 213 was found in an Ashkenazi Jew.

We believe that this is the first report of polymorphic SNPs within a polymorphic *Alu* distribution. This combination of a polymorphic *Alu* insertion and SNPs, both within the insertion and the flanking region, promises to provide an informative X-linked marker for the reconstruction of human population and evolutionary history and in particular the spread of *Alu* repeat elements throughout global populations.

Acknowledgements We would like to thank everyone who kindly provided the DNA samples that made this study possible. This research was supported by grants from the National Science Foundation to MAB (grant # BCS-0218338) and the Canadian Institutes of Health Research to DL (grant # MOP-12782).

References

- Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E, Zuckerkandl E (1996) Standardized nomenclature for *Alu* repeats. *J Mol Evol* 42:3–6
- Blonden LA, Terwindt GM, Den Dunnen JT, Van Ommen GJ (1994) A polymorphic STS in intron 44 of the dystrophin gene. *Hum Genet* 93:479–480
- Dobson-Stone C, Cox RD, Lonie L, Southam L, Fraser M, Wise C, Bernier F, Hodgson S, Porter DE, Simpson AH, Monaco AP (2000) Comparison of fluorescent single-strand conformation polymorphism analysis and denaturing high-performance liquid chromatography for detection of EXT1 and EXT2 mutations in hereditary multiple exostoses. *Eur J Hum Genet* 8:24–32
- Gross E, Arnold N, Goette J, Schwarz-Boeger U, Kiechle M (1999) A comparison of BRCA1 mutation analysis by direct sequencing, SSCP and DHPLC. *Hum Genet* 105:72–78
- Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci USA* 86:2766–2770
- 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
- Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, Carroll ML, Nguyen SV, Walker JA, Prasad BV, 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
- Zietkiewicz E, Richer C, Makalowski W, Jurka J, Labuda D (1994) A young *Alu* subfamily amplified independently in human and African great apes lineages. *Nucleic Acids Res* 22:5608–5612
- Zietkiewicz E, Yotova V, Jarnik M, Korab-Laskowska M, Kidd KK, Modiano D, Scozzari R, Stoneking M, Tishkoff S, Batzer M, Labuda D (1997) Nuclear DNA diversity in worldwide distributed human populations. *Gene* 205:161–171
- Zietkiewicz E, Yotova V, Gehl D, Wambach T, Arrieta I, Batzer M, Cole DEC, Hechtman P, Kaplan F, Modiano D, Moisan JP, Michalski R, Labuda D (2003) Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern human's diversity. *Am J Hum Genet* 73:994–1015