Comprehensive analysis of *Alu*-associated diversity on the human sex chromosomes

Pauline A. Callinan[a,1], Dale J. Hedges[a,1], Abdel-Halim Salem[a,b], Jinchuan Xing[a],
Jerilyn A. Walker[a], Randall K. Garber[a], W. Scott Watkins[c], Michael J. Bamshad[c,d],
Lynn B. Jorde[c], Mark A. Batzer[a,*]

[a] *Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA*
[b] *Department of Anatomy, Faculty of Medicine, Suez Canal University, Ismailia, Egypt*
[c] *Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, UT 84112, USA*
[d] *Department of Pediatrics, University of Utah Health Sciences Center, Salt Lake City, UT 84112, USA*

Abstract

A comprehensive analysis of the human sex chromosomes was undertaken to assess *Alu*-associated human genomic diversity and to identify novel *Alu* insertion polymorphisms for the study of human evolution. Three hundred forty-five recently integrated *Alu* elements from eight different *Alu* subfamilies were identified on the X and Y chromosomes, 225 of which were selected and analyzed by polymerase chain reaction (PCR). From a total of 225 elements analyzed, 16 were found to be polymorphic on the X chromosome and one on the Y chromosome. In line with previous research using other classes of genetic markers, our results indicate reduced *Alu*-associated insertion polymorphism on the human sex chromosomes, presumably reflective of the reduced recombination rates and lower effective population sizes on the sex chromosomes. The *Alu* insertion polymorphisms identified in this study should prove useful for the study of human population genetics.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Retroposon; Insertion polymorphism; Identical by descent

1. Introduction

1.1. Recently integrated Alu insertions in the human genome

Alu elements are a class of repetitive mobile sequences that are dispersed ubiquitously throughout the genomes of primates (Deininger and Batzer, 1993; Schmid, 1996; Batzer and Deininger, 2002). As short interspersed elements (SINEs), *Alu* repeats are the largest family of mobile genetic elements within the human genome, having reached a copy number of over 1 million during the last 65 Myr (million years) (Batzer and Deininger, 2002). *Alu* elements have achieved this copy number by duplicating via an RNA intermediate that is reverse-transcribed by target primed reverse transcription and integrated into the genome (Luan et al., 1993; Kazazian and Moran, 1998). While unable to retropose autonomously, *Alu* elements are thought to appropriate the necessary mobilization machinery from the long interspersed element (LINE) retrotransposon family (Sinnett et al., 1992; Boeke, 1997), which encodes a protein possessing endonuclease and reverse transcriptase activity (Feng et al., 1996; Jurka, 1997).

Phylogenetic studies of *Alu* elements suggest that only a small number of *Alu* elements, deemed ''master'' or source genes, are retropositionally competent (Deininger et al., 1992). Over time, the eventual accumulation of new mutations within these ''master'' or source genes created a hierarchy of *Alu* subfamilies (Deininger et al., 1992; Batzer

---

and Deininger, 2002). Diagnostic mutation sites can be used to classify each individual element according to subfamily and to stratify *Alu* subfamily members based upon age from the oldest (designated J) to intermediate (S) and youngest (Y) (Batzer et al., 1996). Some young *Alu* subfamilies have amplified so recently that they are virtually absent from the genomes of nonhuman primates (Batzer and Deininger, 2002). As a result of the recent integration of young *Alu* subfamily members within the human genome, individual humans can be polymorphic for the presence of *Alu* elements at particular loci. Because the likelihood of two *Alu* elements independently inserting into the same exact location of the genome is extremely small, and as there are no known biological mechanisms for the specific excision of *Alu* elements from the genome, *Alu* insertions can be considered identical by descent or homoplasy-free characters for the study of human population genetics (Batzer and Deininger, 2002; Roy-Engel et al., 2002). SINE insertion polymorphisms are generally thought to be homoplasy-free characters for phylogenetic studies (Shedlock and Okada, 2000; Batzer and Deininger, 2002) and have been utilized to resolve the relationships of artiodactyls and whales (Nikaido et al., 1999, 2001).

1.2. Repetitive elements and genetic variation on the sex chromosomes

The aim of the present study is to annotate young *Alu* insertions on the human sex chromosomes in order to assess *Alu*-associated diversity and to identify new *Alu* insertion polymorphisms. Several previous studies have focused on the evolutionary dynamics of repetitive elements on the sex chromosomes. Increased accumulation of repetitive elements on the X and Y has been detected in humans and other taxa (Wichman et al., 1992; Charlesworth et al., 1994; Smit, 1999; Erlandsson et al., 2000; Boissinot et al., 2001). The differential accumulation of mobile elements is thought to result from reduced recombination and lower effective population sizes in the sex chromosomes leading to increased fixation of slightly deleterious insertions. However, Boissinot et al. (2001) found sex chromosome enrichment for full-length and greater than 500-bp L1 elements, while demonstrating no associated enrichment in SINEs. Their results suggest that, unlike the longer-length L1 mobile elements, *Alu* insertions may not be deleterious enough, on average, to exhibit a sex chromosome distribution bias.

While no previous research specifically addresses repetitive element-generated insertion polymorphisms on the sex chromosomes, studies using other classes of genetic markers have shown reduced genetic variation on the X and Y chromosomes of humans and other organisms (Nachman, 1997; Begun and Whitley, 2000; Yu et al., 2001). This reduction of observed polymorphism has largely been attributed to reduced recombination and lower effective population sizes of these chromosomes (Nachman, 1997; Begun and Whitley, 2000). The current study affords the opportunity to assess human sex chromosome variability with a novel class of genetic markers.

2. Materials and methods

2.1. Cell lines and DNA samples

The DNA samples used in this study were isolated from the cell lines as follows: human (*Homo sapiens*), HeLa (ATCC CCL-2), chimpanzee (*Pan troglodytes*) (NG06939), and lowland gorilla (*Gorilla gorilla*) (NG05251). All nonhuman primate cell lines were obtained from the Coriell Institute for Medical Research (Camden, NJ). Human DNA samples from the African–American, Asian, European, and Egyptians were described previously (Carroll et al., 2001). Indian DNA samples of defined sex were described previously (Bamshad et al., 2001). The South American human DNA samples were part of a human diversity panels (HD 17 and 18) purchased from the Coriell Institute for Medical Research.

2.2. Identification of Alu elements

Alu elements from the recently integrated *Alu* subfamilies Ya5, Ya5a2, Ya8, Yb8, Yb9, Yc1, Yd3, and Yd6 were identified from the August 2001 release of the UC Santa Cruz draft sequence (http://genome.ucsc.edu/). *Alu* subfamily members were located by two complementary methods. A local installation of RepeatMasker (http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker) was used to screen sequences on chromosomes X and Y for the positions of recently integrated *Alu* elements. Exceptions to this were the Yc1 and Yc2 subfamilies, which were not identified by the software at the time of the study. In addition, subfamily specific oligonucleotides (Table 1) were utilized in a local installation of the National Center for Biotechnology Information basic local alignment search tool (BLAST) software (Altschul et al., 1990) to identify exact complements within the draft human genomic sequence as

Table 1

Alu subfamily-specific oligonucleotides[a]

Ya5/Ya5a2	5′-CCATCCCGGCTAAAAC-3′
Ya8	5′-ACTAAAACTACAAAAAATAG-3′
Yb8/Yb9	5′-ACTGCAGTCCGCAGTCCGGCC-3′
Yc1/Yc2	5′-GGGCGTGGTAGCGGGCGCCTG-3′
Yd3/Yd6[b]	5′-CGAGACCACGGTGAAACCCCGTC-3′

[a] Subfamilies Ya5/Ya5a2, Yb8/Yb9, Yd3/Yd6, and Yc1/Yc2 were screened using the same oligonucleotide and subsequently differentiated using multiple alignments and/or RepeatMasker.

[b] The Yd3/Yd6 oligonucleotide listed will match all members of the Yd lineage. Yd3 and Yd6 members are subsequently identified by multiple alignment.

previously described. Results from these analyses were pooled and cross-checked to remove duplicate elements. *Alu* elements were then extracted from their locations within the chromosome and aligned with MEGALIGN (DNASTAR V 3.1.7) for subfamily verification and further analysis. Lists of all the *Alu* elements identified in the database searches and full alignments of all the recovered *Alu* elements are available under the publications section of our website (http://batzerlab.lsu.edu).

2.3. Primer design and amplification

Oligonucleotide primers for the polymerase chain reaction (PCR) amplification of each *Alu* element were designed using the Primer3 program (http://www-genome. wi.mit.edu/cgi-bin/primer/primer3_www.cgi). Sequences flanking the *Alu* insertions were first masked with Repeat-Masker to remove all repetitive elements. Primer3 was then utilized to design PCR primers within the remaining flanking unique DNA sequences. PCR amplification was accomplished in 25-µl reactions using either 60 ng of template DNA (human populations) or 15 ng (nonhuman primates) of 0.2 µM of each oligonucleotide primer, 200 µM deoxynucleotide triphosphates, 1.5 mM $MgCl_2$, 10 mM Tris–HCl (pH 8.4), and Taq® DNA polymerase (1 U). Each sample was subjected to the same amplification cycle as follows: initial denaturation of 150 s at 94 °C, 32 cycles of 1 min of denaturation at 94 °C, 1 min at the specific annealing temperature (see http://batzerlab. lsu.edu), 1 min of extension at 72 °C, followed by a final extension at 72 °C for 10 min. For analysis, 20 µl of the PCR products was fractionated on a 2% agarose gel, which contained 0.25 µg/ml ethidium bromide. PCR products were visualized using ultraviolet (UV) fluorescence. Twenty individuals from four populations (African–American, Asian, European, and either Egyptian or South American) were screened to test each locus for insertion polymorphism. Additional male DNA samples from the following populations: French (8 individuals), Indian (15), and African–American (15) were used to confirm polymorphism on the Y chromosome.

3. Results

3.1. Subfamily copy number and distribution

Following a computational search of the human draft sequence, using both diagnostic oligonucleotide queries of the database and RepeatMasker screening, 345 *Alu* repeat elements from eight young *Alu* subfamilies (*Alu* Ya5, *Alu* Ya8, *Alu* Ya5a2, *Alu* Yb8, *Alu* Yb9, *Alu* Yc1, *Alu* Yd3, and *Alu* Yd6) were identified. Of these, 264 recently integrated *Alu* subfamily members were found on human chromosome X, while chromosome Y contained 80. The expected distributions of young *Alu* subfamilies on the sex chromo-

somes were calculated based on the size of each *Alu* subfamily and the proportion of the human draft sequence represented by the respective chromosome (chromosome sizes and sequenced base pair totals taken from the August 2001 freeze UC Santa Cruz summary statistics) as reported previously for human chromosome 19 (Arcot et al., 1998). The results of the database screening and expected numbers are given in Table 2. While several subfamilies were represented at or near expected levels, some deviated substantially. In particular, the number of *Alu* Ya5 elements was double that expected on the Y chromosome, but nearly equal to that expected on the X. The number of Yb8 subfamily members was consistent with expected numbers on both sex chromosomes. The Yc1 subfamily had approximately twice the expected number of elements on both the X and Y chromosomes. However, the excess of Yc1 *Alu* elements probably reflects the erroneous detection of Y subfamily elements that have had a fortuitous single base-pair mutation to the Yc1 consensus sequence (Roy-Engel et al., 2001).

3.2. Age of Alu insertions on the sex chromosomes

The average ages of the recently integrated *Alu* insertions on the X and Y chromosomes were estimated and compared to previous subfamily age estimates to determine if the amplification dynamics of recently integrated *Alu* elements on the sex chromosomes is comparable to that of the rest of the nuclear genome. In order to estimate the average age for each *Alu* subfamily, the number of substitutions at CpG and non-CpG sites was determined. The mutation density for each of these mutation classes is different as a result of the methylation and subsequent spontaneous deamination of 5 methyl-cytosine bases (Bird, 1980) and is approximately 10-fold higher in CpG than non-CpG base positions within *Alu* elements (Labuda and Striker, 1989; Batzer et al., 1990). The average age for each *Alu* subfamily is then estimated by using the mutation

Table 2
Expected and observed distribution of recently integrated *Alu* elements on the X and Y chromosomes

Alu subfamily	Genomic copies[a]	Expected on X[b]	Found on X	Expected on Y[b]	Found on Y
Ya5	2640	130.15	119	20.59	45
Ya8	60	2.96	0	0.47	2
Ya5a2	35	1.73	1	0.27	1
Yb8	1852	91.30	91	14.45	19
Yc1	381	18.78	37	2.97	10
Yb9	79	3.89	7	0.62	1
Yd3	198	9.76	7	1.54	0
Yd6	97	4.78	2	0.76	2

[a] Copy numbers based on previous estimated size of the subfamilies (Batzer and Deininger, 2002; Xing et al., 2003).
[b] Expected number estimated based on the subfamily size and amount of X or Y chromosome sequence in the database as outlined in the text.

density and a neutral rate of evolution of 0.15% per million years for non-CpG sequences (Miyamoto et al., 1987) and 1.5% per million years for CpG sequences as described previously. All deletions, insertions, simple sequence repeat expansions, and truncations were eliminated from the age calculations. All of the *Alu* elements that were identified in the draft sequence and were less than 100 bp in length were eliminated from the analysis. The estimated ages of Ya5, Yb8, and Yc1 are in line with the age estimates that were reported previously (Carroll et al., 2001; Roy-Engel et al., 2001; Xing et al., 2003) of 2.1–4.2 Myr and are summarized in Table 3. Subfamilies with less than five representatives on the sex chromosomes were excluded as there was not enough sequence for accurate estimates to be made. It is important to note that the mutation rate for X and Y chromosome DNA sequences is different (Huang et al., 1997), and these differences may influence these age estimates. However, this difference should be minimal.

An evolutionary analysis of the time of origin of the *Alu* elements located on the human sex chromosomes was determined within the primate lineage by PCR amplification of the individual loci using chimpanzee and gorilla DNA as templates. From the 225 recently integrated *Alu* elements analyzed in this study, three X chromosome loci (Yc1DP26, Yc1DP8, and Ya5DP38) and three Y chromosome loci (Yc1AD168, Yc1AD242, and Yc1AD244) contained insertions within the chimpanzee and/or gorilla genomes, confirming that the overwhelming majority of the sex chromosome-specific *Alu* elements inserted in the human genome after the human and African ape divergence, which is thought to have occurred within the last 4–6 Myr. It is interesting to note that most of the putative recently integrated *Alu* elements that were also found in nonhuman primate genomes were members of the Yc1 family. This is not surprising since a single base mutation differentiates this subfamily from the *Alu* Y subfamily as mentioned above (Roy-Engel et al., 2001).

3.3. Human genomic diversity

Individual *Alu* elements were screened for polymorphism by amplification of a panel of diverse human DNA samples, which included 20 African–Americans, 20 Europeans, 20 Asians, and either 20 Egyptians or South Americans. A total of 80 individuals were screened, comprising approximately 120 X chromosomes and 40 Y chromosomes (Table 4). One hundred twenty-one sex chromosome-specific *Alu* elements were not amplified by PCR, 109 of which were positioned within repeat-saturated regions of the genome, making the design of unique primers impossible. The remaining 12 elements either generated paralogous PCR products, or failed to amplify for unknown reasons, which may include mutations within the sites where the oligonucleotide primers anneal and small deletions or even larger recombination events between adjacent sequences such as mobile elements.

The number of elements on the X chromosome, which exhibited polymorphism within the human genomes that were surveyed, consisted of nine Ya5′s, five Yb8′s, one Ya5a2, and one Yd3 element. All young subfamily members analyzed on the Y chromosome were found to be monomorphic, with the exception of one previously identified Yb8 *Alu* insertion, termed YAP (Y *Alu* polymorphism) (Hammer, 1994), which is an intermediate frequency *Alu* insertion polymorphism. The remaining *Alu* insertion polymorphisms were classified as high, low, or intermediate frequency as previously described and summarized in Table 4. Unbiased heterozygosity values for each of the polymorphisms were determined by allele counting. The heterozygosity data suggest that the *Alu* insertion polymorphisms from the X chromosome will be useful as genetic markers for human population genetics. A schematic diagram showing the location of all the *Alu* insertion polymorphisms located on the human X and Y chromosomes is shown in Fig. 1.

The levels of *Alu* insertion polymorphism on the X and Y chromosomes were compared to previous data on the detection of autosomal *Alu* insertion polymorphisms. The data in Carroll et al. (2001) were adapted to exclude all elements on the sex chromosomes in order to make comparisons against autosomal loci only. Chromosome X showed 14.06% (9/64) polymorphism, for the Ya5 subfamily, 100% (1/1) for Ya5a2, 20% (1/5) for the Yd3 subfamily and 8.77% (5/57) for the Yb8 subfamily. On the Y chromosome, 6.66% (3/45) polymorphism was observed for the Ya5 subfamily, 10.53% (2/19) for the Yb8 subfamily, and 50% (1/2) for the Yb9 subfamily. Compared to previously reported levels of *Alu* insertion polymorphism throughout the genome of 25% (Ya5), 80% (Ya5a2), 20% (Yb8), and 25% (Yc1) (Batzer and Deininger, 2002), our data indicate

Table 3
Estimated ages of sex chromosome-specific *Alu* subfamilies

Alu subfamily	Ya5		Yb8		Yc1		Yd3	
Chromosome	X	Y	X	Y	X	Y	X	Y
Number of loci analyzed	119	36	88	17	32	10	7	0
CpG mutation density (%)	2.53	1.97	3.60	1.74	2.5	2.65	12.1	N/A
Non-CpG mutation density (%)	0.78	0.49	0.53	0.47	0.28	0.24	1.39	N/A
Estimated age from CpG mutations (Myr)	1.73	1.35	2.47	1.19	1.72	1.81	6.60	N/A
Estimated age from non-CpG mutations (Myr)	4.92	3.24	3.54	3.16	1.86	1.62	8.03	N/A
Variance (between age estimates) (Myr)	5.09	1.77	5.79	1.94	0.01	0.02	1.37	N/A

Table 4
X chromosome *Alu* insertion polymorphism, genotypes, and heterozygosity

Name	African–American							Asian							European							Egyptian							Average Het[a]
	Genotypes					*fAlu*	Het[b]	Genotypes					*fAlu*	Het[b]	Genotypes					*fAlu*	Het[b]	Genotypes					*fAlu*	Het[b]	
	Female			Male				Female			Male				Female			Male				Female			Male				
	+/+	+/−	−/−	+	−			+/+	+/−	−/−	+	−			+/+	+/−	−/−	+	−			+/+	+/−	−/−	+	−			
(A) Intermediate frequency																													
Ya5a2DP1	2	0	4	3	7	0.32	0.47	3	0	3	10	1	0.37	0.45	0	1	4	1	12	0.09	0.18	6	1	1	8	0	0.09	0.18	0.32
Yb8DP2	5	2	0	9	3	0.81	0.34	0	3	8	1	8	0.13	0.23	0	3	9	1	7	0.13	0.23	2	4	6	2	6	0.31	0.43	0.31
Yd3JX437	1	2	4	5	0	0.33	0.48	3	6	2	6	0	0.58	0.50	0	2	10	0	8	0.07	0.08	0	5	8	1	6	0.18	0.29	0.34
Yb8NBC634	4	2	1	9	0	0.93	0.26	7	0	0	7	0	1.00	0	7	0	0	5	0	1.00	0	7	0	0	10	0	1.00	0	0.07
(B) High frequency																													
Ya5DP57	3	0	4	1	10	0.28	0.41	5	2	0	11	2	0.85	0.27	3	2	0	13	2	0.84	0.31	8	1	0	9	0	0.96	0.06	0.26
Ya5DP62	5	2	0	7	5	0.73	0.43	7	0	0	12	1	0.96	0.08	4	0	0	8	5	0.76	0.36	5	4	0	6	2	0.77	0.38	0.31
Ya5DP77	2	3	2	4	9	0.41	0.52	2	4	0	11	3	0.73	0.43	5	0	0	15	0	1.00	0	5	2	0	9	1	0.88	0.23	0.30
Ya5NBC98	5	2	0	8	5	0.74	0.42	7	0	0	12	1	0.96	0.08	5	1	0	6	6	0.71	0.45	5	4	0	5	1	0.79	0.33	0.32
Ya5NCB491	3	0	4	6	3	0.52	0.53	6	0	1	10	0	0.92	0.14	5	0	0	12	0	1.00	0	10	0	0	7	0	1.00	0	0.17
Yb8DP49	6	1	0	9	3	0.78	0.38	8	3	0	9	0	0.90	0.13	8	4	0	7	1	0.85	0.26	10	2	1	7	0	0.94	0.08	0.21
Yb8NBC102	7	1	0	10	3	0.86	0.27	7	0	0	13	0	1.00	0	5	0	0	15	9	0.74	0.34	10	0	0	10	0	1.00	0	0.15
Yb8NBC578	3	4	0	8	5	0.67	0.48	6	0	0	11	2	0.92	0.16	5	0	0	15	0	1.00	0	10	0	0	6	1	0.96	0.14	0.19
(C) Low frequency																													
Ya5DP3	0	2	4	3	10	0.20	0.35	0	4	3	6	7	0.37	0.50	0	1	4	1	12	0.09	0.18	0	0	8	2	4	0.09	0.30	0.33
Ya5DP4	0	1	6	3	10	0.15	0.28	0	0	6	0	13	0	0	0	0	5	1	11	0.05	0.09	0	2	7	0	6	0.08	0.11	0.12
Ya5NDP13	7	0	0	12	1	0.96	0.08	7	0	0	13	0	1.00	0	5	0	0	15	0	1.00	0	9	0	0	10	0	1.00	0	0.02
Ya5NBC37	2	3	2	4	9	0.41	0.52	2	2	3	5	8	0.41	0.52	0	3	1	3	13	0.25	0.46	0	3	6	0	7	0.12	0.16	0.42

The level of insertion polymorphism was determined as: low frequency—the absence of the element from all individuals tested, except one or two homozygous or heterozygous individuals; intermediate frequency—the *Alu* element is variable as to its presence or absence in at least one population; and high frequency—the element is present in all individuals in all populations tested, except for one or heterozygous individuals.

[a] Average heterozygosity is the average of the population heterozygosity across all four populations.

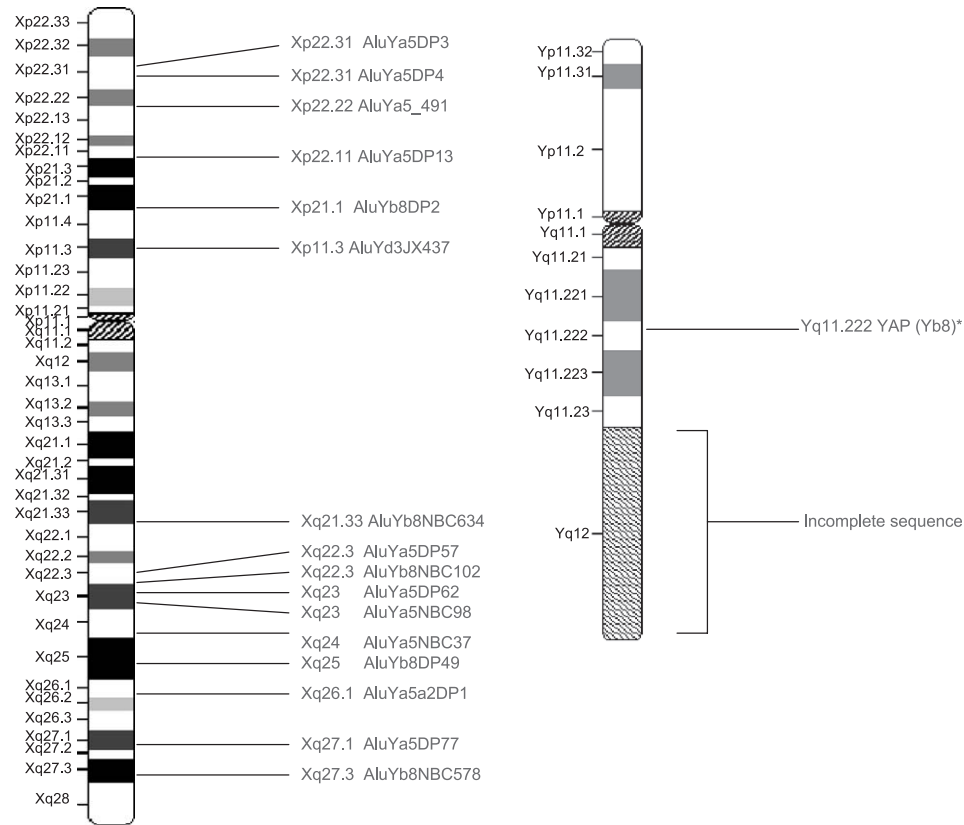[b] This is the unbiased heterozygosity, which takes into account sex differences within the calculation.

Fig. 1. Idiogram of human sex chromosome-specific *Alu* insertion polymorphisms. The physical location of each *Alu* insertion polymorphism was determined using the sequence map from each chromosome as a framework to localize the elements. The sequence from the q12 portion of the human Y chromosome has not yet been completed and therefore the *Alu* elements within this portion of the Y chromosome have not yet been analyzed. All of the *Alu* insertion polymorphisms from the recently integrated subfamilies of elements are shown in the figure. *Denotes the previously reported YAP *Alu* element (Hammer, 1994).

that there is a slight reduction in *Alu* insertion polymorphism on the human sex chromosomes.

4. Discussion

4.1. Distribution of Alu elements

The expected chromosomal distribution of recently integrated *Alu* elements was calculated based on the estimated subfamily size and the relative percentage of the draft sequence constituted by each chromosome. The distribution bias in the observed numbers of *Alu* elements appears to be subfamily-specific and is in good agreement with a recently published analysis sex chromosome mobile elements (Jurka et al., 2002). For example, the Ya5 subfamily has approximately twice the number of *Alu* elements expected on the Y chromosome but nearly equal the number expected on the X chromosome. In contrast, the distribution of Yb8 subfamily members was consistent with estimated expectations on both chromosomes. Population genetics theory predicts that smaller effective populations should result in more frequent fixation of slightly deleterious insertions. Similarly, the virtual lack of recombination on the Y and reduced recombination on the X increases the extent of background

selection and selective sweeps, further lowering the effective population size. Previous studies have reported a higher percentage of repetitive elements on the Y chromosome relative to autosomes and the X chromosome (Erlandsson et al., 2000). Boissinot et al. (2001) previously reported an overrepresentation of full length and >500-bp LINE elements, but no enrichment of SINEs on the sex chromosomes. In addition, the mobilization of *Alu* repeats has recently been suggested to be male germline-specific (Jurka et al., 2002), suggesting yet another mechanism for the differential accumulation of *Alu* repeats within the human genome. Therefore, we conclude that the distribution of different classes of mobile elements on the sex chromosomes in different species is the result of a number of complex processes such as mobilization mechanism and integration site preferences that are mobile element-specific.

4.2. Age of Alu subfamily members

The ages of recently integrated *Alu* elements on the sex chromosomes was estimated based upon CpG and non-CpG mutation densities as reported previously. The estimated ages for the sex chromosome-specific *Alu* elements are in good agreement with those reported previously (Carroll et al., 2001; Roy-Engel et al., 2001). It is possible that the

higher mutation rate in the male germline (Huang et al., 1997) would result in increased divergence and therefore higher estimated ages for *Alu* subfamily members on the Y chromosome. This effect, however, may be more detectable in older *Alu* subfamilies, which have had more time to acquire mutations than in the recently integrated *Alu* subfamilies and certainly should not act selectively upon a single family of elements. This is in good agreement with a previous computational analysis of Y chromosome-specific mobile elements, which demonstrated that the older *Alu* J and *Alu* S subfamilies showed significantly higher divergence on the Y chromosome, while the younger *Alu* Y subfamily divergence did not exhibit a significant difference (Erlandsson et al., 2000). Similarly, due to the increased male mutation rate, X-linked loci should theoretically exhibit a lower mutation rate than their autosomal counterparts since only one out of three X chromosomes is transmitted through the male germline in each generation. However, this effect is likely minimal and is not reflected in the ages of the young *Alu* elements.

4.3. Population dynamics

The recently integrated *Alu* subfamily members on the X and Y chromosomes exhibited reduced polymorphism as compared to their autosomal counterparts. Age estimates and data from orthologous inserts in nonhuman primates indicate that this reduction in polymorphism is not the result of increased age of *Alu* insertions found on the sex chromosomes. Rather, the results are consistent with neutral theory, given that lower effective population size should result in more rapid fixation of elements, lowering overall polymorphism levels on the sex chromosomes. Reduced recombination on the X and Y chromosomes may exacerbate this effect by increasing the extent of background selection and selective sweeps that further remove polymorphism (Charlesworth et al., 1994; Lander et al., 2001). The current findings are in agreement with several previously published studies in humans and other organisms that have found reduced polymorphism on the sex chromosomes (Hammer, 1994; Jorde et al., 2000; Yu et al., 2001).

Aside from the previously identified YAP *Alu* element, all of the *Alu* loci located in the nonrecombining portion of the Y chromosome were monomorphic for the presence of the *Alu* repeat in diverse populations. This suggests that the *Alu*-associated variation currently on the human Y chromosome is very low, probably existing as low-frequency insertions that were not detected in this study, as the young *Alu* elements were ascertained from a single genome and have to be present within that genome in order to be identified. Thus, our data point to an evolutionarily recent event, which dramatically reduced *Alu*-associated Y chromosome diversity, or to an effective population size for the human Y chromosome, which has not been large enough to harbor appreciable *Alu* insertion polymorphism.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Arcot, S.S., Adamson, A.W., Risch, G.W., LaFleur, J., Robichaux, M.B., Lamerdin, J.E., Carrano, A.V., Batzer, M.A., 1998. High-resolution cartography of recently integrated human chromosome 19-specific Alu fossils. J. Mol. Biol. 281, 843–856.

Bamshad, M., Kivisild, T., Watkins, W.S., Dixon, M.E., Ricker, C.E., Rao, B.B., Naidu, J.M., Prasad, B.V., Reddy, P.G., Rasanayagam, A., Papiha, S.S., Villems, R., Redd, A.J., Hammer, M.F., Nguyen, S.V., Carroll, M.L., Batzer, M.A., Jorde, L.B., 2001. Genetic evidence on the origins of Indian caste populations. Genome Res. 11, 994–1004.

Batzer, M.A., Deininger, P.L., 2002. Alu repeats and human genomic diversity. Nat. Rev. Genet. 3, 370–379.

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

Batzer, M.A., Deininger, P.L., Hellmann-Blumberg, U., Jurka, J., Labuda, D., Rubin, C.M., Schmid, C.W., Zietkiewicz, E., Zuckerkandl, E., 1996. Standardized nomenclature for Alu repeats. J. Mol. Evol. 42, 3–6.

Begun, D.J., Whitley, P., 2000. Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. Proc. Natl. Acad. Sci. U. S. A. 97, 5960–5965.

Bird, A.P., 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 8, 1499–1504.

Boeke, J.D., 1997. LINEs and Alus—the polyA connection. Nat. Genet. 16, 6–7.

Boissinot, S., Entezam, A., Furano, A.V., 2001. Selection against deleterious LINE-1-containing loci in the human lineage. Mol. Biol. Evol. 18, 926–935.

Carroll, M.L., Roy-Engel, A.M., Nguyen, S.V., Salem, A.H., Vogel, E., Vincent, B., Myers, J., Ahmad, Z., Nguyen, L., Sammarco, M., Watkins, W.S., Henke, J., Makalowski, W., Jorde, L.B., Deininger, P.L., Batzer, M.A., 2001. Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. J. Mol. Biol. 311, 17–40.

Charlesworth, B., Sniegowski, P., Stephan, W., 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371, 215–220.

Deininger, P.L., Batzer, M.A., 1993. Evolution of retroposons. Evol. Biol. 27, 157–196.

Deininger, P.L., Batzer, M.A., Hutchison III, C.A., Edgell, M.H., 1992. Master genes in mammalian repetitive DNA amplification. Trends Genet. 8, 307–311.

Erlandsson, R., Wilson, J.F., Paabo, S., 2000. Sex chromosomal transposable element accumulation and male-driven substitutional evolution in humans. Mol. Biol. Evol. 17, 804–812.

Feng, Q., Moran, J.V., Kazazian Jr., H.H., Boeke, J.D., 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87, 905–916.

Hammer, M.F., 1994. A recent insertion of an alu element on the Y chromosome is a useful marker for human population studies. Mol. Biol. Evol. 11, 749–761.

Huang, W., Chang, B.H., Gu, X., Hewett-Emmett, D., Li, W., 1997. Sex

differences in mutation rate in higher primates estimated from AMG intron sequences. J. Mol. Evol. 44, 463–465.

Jorde, L.B., Watkins, W.S., Bamshad, M.J., Dixon, M.E., Ricker, C.E., Seielstad, M.T., Batzer, M.A., 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am. J. Hum. Genet. 66, 979–988.

Jurka, J., 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc. Natl. Acad. Sci. U. S. A. 94, 1872–1877.

Jurka, J., Krnjaic, M., Kapitonov, V., Stenger, J.E., Kokhanyy, O., 2002. Active Alu elements are passed primarily through paternal germ lines. Theor. Popul. Biol.

Kazazian Jr., H.H., Moran, J.V., 1998. The impact of L1 retrotransposons on the human genome. Nat. Genet. 19, 19–24.

Labuda, D., Striker, G., 1989. Sequence conservation in Alu evolution. Nucleic Acids Res. 17, 2477–2491.

Lander, E.S., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409 (6822), 860–921.

Luan, D.D., Korman, M.H., Jakubczak, J.L., Eickbush, T.H., 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72, 595–605.

Miyamoto, M.M., Slightom, J.L., Goodman, M., 1987. Phylogenetic relations of humans and African apes from DNA sequences in the psi eta-globin region. Science 238, 369–373.

Nachman, M.W., 1997. Patterns of DNA variability at X-linked loci in *Mus domesticus*. Genetics 147, 1303–1316.

Nikaido, M., Rooney, A.P., Okada, N., 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. U. S. A. 96, 10261–10266.

Nikaido, M., Matsuno, F., Hamilton, H., Brownell Jr., R.L., Cao, Y., Ding, W., Zuoyan, Z., Shedlock, A.M., Fordyce, R.E., Hasegawa, M., Okada, N., 2001. Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. Proc. Natl. Acad. Sci. U. S. A. 98, 7384–7389.

Roy-Engel, A.M., Carroll, M.L., Vogel, E., Garber, R.K., Nguyen, S.V., Salem, A.H., Batzer, M.A., Deininger, P.L., 2001. Alu insertion polymorphisms for the study of human genomic diversity. Genetics 159, 279–290.

Roy-Engel, A.M., Carroll, M.L., El-Sawy, M., Salem, A.H., Garber, R.K., Nguyen, S.V., Deininger, P.L., Batzer, M.A., 2002. Non-traditional Alu evolution and primate genomic diversity. J. Mol. Biol. 316, 1033–1040.

Schmid, C.W., 1996. Alu: structure, origin, evolution, significance and function of one-tenth of human DNA. Prog. Nucleic Acid Res. Mol. Biol. 53, 283–319.

Shedlock, A.M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. Bioessays 22, 148–160.

Sinnett, D., Richer, C., Deragon, J.M., Labuda, D., 1992. Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences. J. Mol. Biol. 226, 689–706.

Smit, A.F., 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr. Opin. Genet. Dev. 9, 657–663.

Wichman, H.A., Van den Bussche, R.A., Hamilton, M.J., Baker, R.J., 1992. Transposable elements and the evolution of genome organization in mammals. Genetica 86, 287–293.

Xing, J., Salem, A.-H., Hedges, D.J., Kilroy, G.E., Watkins, W.S., Schienman, J.E., Stewart, C.-B., Jurka, J., Jorde, L.B., Batzer, M.A., 2003. Comprehensive analysis of two Alu Yd subfamilies. J. Mol. Evol., in press.

Yu, N., Zhao, Z., Fu, Y.X., Sambuughin, N., Ramsay, M., Jenkins, T., Leskinen, E., Patthy, L., Jorde, L.B., Kuromori, T., Li, W.H., 2001. Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. Mol. Biol. Evol. 18, 214–222.