Comprehensive analysis of \textit{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$^a, *$

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

Received 8 August 2002; received in revised form 17 September 2002; accepted 27 October 2002

Abstract

A comprehensive analysis of the human sex chromosomes was undertaken to assess \textit{Alu}-associated human genomic diversity and to identify novel \textit{Alu} insertion polymorphisms for the study of human evolution. Three hundred forty-five recently integrated \textit{Alu} elements from eight different \textit{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 \textit{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 \textit{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 \textit{Alu} insertions in the human genome

\textit{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), \textit{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). \textit{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, \textit{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 \textit{Alu} elements suggest that only a small number of \textit{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 \textit{Alu} subfamilies (Deininger et al., 1992; Batzer...
1.2. Repeat 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; Smidt, 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 Whitely, 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 Whitely, 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 non-human 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>
<thead>
<tr>
<th>Table 1 Alu subfamily-specific oligonucleotidesa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ya5/Ya5a2 5'-CCATCCCCGCTAAAAC-3'</td>
</tr>
<tr>
<td>Ya8     5'-ACTACCCCATACCA-3'</td>
</tr>
<tr>
<td>Yb8/Yb9 5'-ACTGCCATGCCAGTCCGCC-3'</td>
</tr>
<tr>
<td>Yc1/Yc2 5'-GGGCCGTGTAGCCGCGCCCT-3'</td>
</tr>
<tr>
<td>Yd3/Yd65 5'-CGAGACCCAGGTTGAAACCCCGTC-3'</td>
</tr>
</tbody>
</table>

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 RepeatMasker 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 MgCl₂, 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 chromosomes 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>
<thead>
<tr>
<th><em>Alu</em> subfamily</th>
<th>Genomic copies[^a]</th>
<th>Expected on X[^b]</th>
<th>Found on X</th>
<th>Expected on Y[^b]</th>
<th>Found on Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ya5</td>
<td>2640</td>
<td>130.15</td>
<td>119</td>
<td>20.59</td>
<td>45</td>
</tr>
<tr>
<td>Ya8</td>
<td>60</td>
<td>2.96</td>
<td>0</td>
<td>0.47</td>
<td>2</td>
</tr>
<tr>
<td>Ya5a2</td>
<td>35</td>
<td>1.73</td>
<td>1</td>
<td>0.27</td>
<td>1</td>
</tr>
<tr>
<td>Yb8</td>
<td>1852</td>
<td>91.30</td>
<td>91</td>
<td>14.45</td>
<td>19</td>
</tr>
<tr>
<td>Yc1</td>
<td>381</td>
<td>18.78</td>
<td>37</td>
<td>2.97</td>
<td>10</td>
</tr>
<tr>
<td>Yb9</td>
<td>79</td>
<td>3.89</td>
<td>7</td>
<td>0.62</td>
<td>1</td>
</tr>
<tr>
<td>Yd3</td>
<td>198</td>
<td>9.76</td>
<td>7</td>
<td>1.54</td>
<td>0</td>
</tr>
<tr>
<td>Yd6</td>
<td>97</td>
<td>4.78</td>
<td>2</td>
<td>0.76</td>
<td>2</td>
</tr>
</tbody>
</table>

[^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

<table>
<thead>
<tr>
<th>Alu subfamily</th>
<th>Ya5</th>
<th>Yb8</th>
<th>Yc1</th>
<th>Yd3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome</td>
<td>X</td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Number of loci analyzed</td>
<td>119</td>
<td>36</td>
<td>88</td>
<td>17</td>
</tr>
<tr>
<td>CpG mutation density (%)</td>
<td>2.53</td>
<td>1.97</td>
<td>3.60</td>
<td>1.74</td>
</tr>
<tr>
<td>Non-CpG mutation density (%)</td>
<td>0.78</td>
<td>0.49</td>
<td>0.53</td>
<td>0.47</td>
</tr>
<tr>
<td>Estimated age from CpG mutations (Myr)</td>
<td>1.73</td>
<td>1.35</td>
<td>2.47</td>
<td>1.19</td>
</tr>
<tr>
<td>Estimated age from non-CpG mutations (Myr)</td>
<td>4.92</td>
<td>3.24</td>
<td>3.54</td>
<td>3.16</td>
</tr>
<tr>
<td>Variance (between age estimates) (Myr)</td>
<td>5.09</td>
<td>1.77</td>
<td>5.79</td>
<td>1.94</td>
</tr>
</tbody>
</table>
Table 4
X chromosome Alu insertion polymorphism, genotypes, and heterozygosity

<table>
<thead>
<tr>
<th>Name</th>
<th>African–American</th>
<th>Asian</th>
<th>European</th>
<th>Egyptian</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Genotypes</td>
<td></td>
<td></td>
<td></td>
<td>Hetb</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>fAlu</td>
<td>Hetb</td>
<td>fAlu</td>
<td>Hetb</td>
<td>fAlu</td>
</tr>
<tr>
<td></td>
<td>+/+</td>
<td>+/−</td>
<td>+/+</td>
<td>+/−</td>
<td>+/+</td>
</tr>
</tbody>
</table>
| (A) Intermediate frequency
| Ya5a2DP1      | 2     | 4    | 3     | 7    | 0.32   | 0.47  | 3     | 0    | 3     | 10    | 0.37   | 0.45  | 0     | 1    | 4    | 12    | 0.09  | 0.18  | 6     | 1     | 1     | 8     | 0.09  | 0.18  | 0.32  |
| Yb8DP2        | 5     | 2    | 0     | 9    | 3     | 0.81  | 0.34  | 0     | 3    | 8     | 1     | 0.13   | 0.23  | 0     | 3    | 9    | 1     | 7     | 0.13  | 0.23  | 2     | 4     | 6     | 2     | 6.31  | 0.43  | 0.31  |
| Yd3X437       | 1     | 2    | 4     | 5    | 0     | 0.33  | 0.48  | 3     | 6    | 2     | 6     | 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     | 0.96  | 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  | 0.17  | 0     |
| 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  | 0.15  | 0.17  |
| 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  |

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 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.
that there is a slight reduction in \textit{Alu} insertion polymorphism on the human sex chromosomes.

4. Discussion

4.1. Distribution of \textit{Alu} elements

The expected chromosomal distribution of recently integrated \textit{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 \textit{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 \textit{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 \textit{Alu} repeats has recently been suggested to be male germline-specific (Jurka et al., 2002), suggesting yet another mechanism for the differential accumulation of \textit{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 \textit{Alu} subfamily members

The ages of recently integrated \textit{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 \textit{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.

Acknowledgements

This research was supported by the National Institutes of Health (RO1 GM59290; L.B.J. and M.A.B.) and the Louisiana Board of Regents Millennium Trust Health Excellence Fund (HEF) [(2000-05)-05, (2000-05)-01, and (2001-06)-02 (M.A.B.)].

References

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


