JMB

Large-scale Analysis of the Alu Ya5 and Yb8 Subfamilies and their Contribution to Human Genomic Diversity

Marion L. Carroll[1], Astrid M. Roy-Engel[2], Son V. Nguyen[1]
Abdel-Halim Salem[1,3], Erika Vogel[2], Bethaney Vincent[1,3]
Jeremy Myers[1,3] Zahid Ahmad[1,3], Lan Nguyen[1], Mimi Sammarco[1]
W. Scott Watkins[4] Jurgen Henke[5], Wojciech Makalowski[6]
Lynn B. Jorde[4] Prescott L. Deininger[2,7] and Mark A. Batzer[1,3]*

[1]*Departments of Pathology, Genetics Biochemistry and Molecular Biology Stanley S. Scott Cancer Center Neuroscience Center of Excellence Louisiana State University Health Sciences Center, 1901 Perdido Street New Orleans, LA 70112, USA*

[2]*Tulane Cancer Center, Department of Environmental Health Sciences Tulane University Health Sciences Center, 1430 Tulane Ave., SL-66 New Orleans, LA 70112, USA*

[3]*Department of Biological Sciences Biological Computation and Visualization Center, Louisiana State University, 508 Life Sciences Building, Baton Rouge, LA 70803, USA*

[4]*Department of Human Genetics University of Utah Health Sciences Center, Salt Lake City, UT 84112, USA*

[5]*Institut fur Blutgruppenforschung Hohenzollernring 57, D-50501, Koln Germany*

[6]*National Center of Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA*

[7]*Laboratory of Molecular Genetics Alton Ochsner Medical Foundation 1516 Jefferson Highway, New Orleans LA 70121, USA*

*Corresponding author

We have utilized computational biology to screen GenBank for the presence of recently integrated Ya5 and Yb8 Alu family members. Our analysis identified 2640 Ya5 Alu family members and 1852 Yb8 Alu family members from the draft sequence of the human genome. We selected a set of 475 of these elements for detailed analyses. Analysis of the DNA sequences from the individual Alu elements revealed a low level of random mutations within both subfamilies consistent with the recent origin of these elements within the human genome. Polymerase chain reaction assays were used to determine the phylogenetic distribution and human genomic variation associated with each Alu repeat. Over 99 % of the Ya5 and Yb8 Alu family members were restricted to the human genome and absent from ortholo-gous positions within the genomes of several non-human primates, confirming the recent origin of these Alu subfamilies in the human genome. Approximately 1 % of the analyzed Ya5 and Yb8 Alu family members had integrated into previously undefined repeated regions of the human genome. Analysis of mosaic Yb8 elements suggests gene conversion played an important role in generating sequence diversity among these elements. Of the 475 evaluated elements, a total of 106 of the Ya5 and Yb8 Alu family members were polymorphic for inser-tion presence/absence within the genomes of a diverse array of human populations. The newly identified Alu insertion poly-morphisms will be useful tools for the study of human genomic diversity.

© 2001 Academic Press

Keywords: Alu insertion polymorphism; gene conversion; computational biology

Introduction

Alu elements are the most abundant Short INterspersed Elements (SINEs), reaching a copy number of over one million in the human genome,[1] making them the mobile element with the highest copy number. Alu repeats compose greater than 10% of the mass of the human genome. Full-length Alu elements are approximately 300 bp in length and commonly found in introns, 3′ untranslated regions of genes, and intergenic genomic regions.[2−4] Amplification of Alu elements occurs through the reverse transcription of RNA in a process termed retroposition.[5] However, Alu elements have no open reading frames, so they are thought to parasitize the required factors for their amplification from Long Interspersed Elements (LINEs).[6−8] Although the human genome contains over one million Alu elements, only a few Alu elements, termed "master" or source genes, are retroposition competent.[9−13] The crucial factor(s) that determine an Alu as a functional source gene are not fully known. Several factors have been suggested to influence the amplification process, including transcriptional capacity, priming or self-priming for reverse transcription and others.[14]

Alu elements first appeared in the primate genomes over 65 million years (myr) ago.[11] Since then, the amplification of Alu elements within the human genome has been punctuated, with the current rate being at least 100-fold slower than the initial rate of Alu expansion within primate genomes.[15] Throughout Alu evolution, the source gene(s) accumulated mutations that were incorporated into the new copies made, creating new Alu subfamilies. Therefore, the Alu family is composed of a number of distinct subfamilies characterized by a hierarchical series of mutations that result in a series of subfamilies of different ages.[15−20] Of these subfamilies, almost all of the recently integrated Alu elements within the human genome belong to one of several closely related "young" Alu subfamilies: Y, Yc1, Yc2, Ya5, Ya5a2, Ya8, Yb8, and Yb9 with the majority being Ya5 and Yb8 subfamily members.[9,18,21,22]

The availability of a draft human genomic DNA sequence as a result of the Human Genome Project[23] facilitates the *"in silico"* identification of recently integrated Alu elements from the human genome.[17,18] This method proves to be less demanding in comparison to older approaches, such as cloning and library screening.[9,21,24] These recently integrated Alu elements serve as temporal landmarks in the evolution of our genome, and many of them will prove to be useful in the study of human evolution and in the study of the natural history of different regions of the genome. Here, we present an analysis of the human genomic diversity associated with 475 members of the Alu Ya5 and Yb8 subfamilies in the human genome.

Results

Subfamily copy number and sequence diversity

In order to determine the copy number of each subfamily of Alu elements, we searched the draft sequence of the entire human genome for the presence of Alu repeats using oligonucleotide sequences complementary to each of the subfamilies (outlined in the Materials and Methods). Our query of the draft human genome sequence identified 2640 Alu Ya5 subfamily members and 1852 Alu Yb8 subfamily members. Both of these copy numbers are in good agreement with previous estimates of the sizes of these Alu subfamilies based upon high-resolution restriction mapping and computational biology.[18,21]

A comparison of the nucleotide sequences of all of the Ya5 and Yb8 Alu family members can be found at our website (http://129.81.225.52). In order to determine the time of origin for the respective Ya5 and Yb8 subfamilies, we divided the nucleotide substitutions within the elements in each family into those that occurred in CpG dinucleotides and those that occurred in non-CpG positions. The distinction between types of mutations is made because the CpG dinucleotides mutate at a rate that is about ten times faster than non-CpG positions[9,25] as a result of the deamination of 5-methylcytosine.[26] In addition, all insertions, deletions and 5′ truncations were excluded from our calculations. A total of 441 non-CpG and 241 CpG mutations occurred within the 231 Alu Ya5 subfamily members used in this analysis. For the 244 Alu Yb8 subfamily members analyzed, a total of 478 non-CpG and 275 CpG mutations were observed. Using a neutral rate of evolution for primate intervening DNA sequences of 0.15% per million years[27] and the non-CpG mutation density of 0.799% (441/55,209) within the 231 Ya5 Alu elements yields an estimated age of 5.32 million years for the Ya5 subfamily members. Using only non-CpG mutations in the 244 Yb8 sequences yields an estimate of 5.30 million years old for the Yb8 subfamily (478/60,024). This estimate of age is somewhat higher than the 2.7-4.1 million years previously reported.[21] However, the previous study of Ya5 and Yb8 Alu family members involved only a small number of elements making the calculated subfamily ages more subject to random statistical fluctuation. Alternatively, the new estimated age based upon non-CpG mutations may be artificially inflated due to sequencing errors in the human draft sequence that may account for an increase in the number of mutations observed.

We can also estimate the ages of each Alu subfamily using CpG-based mutations. The only difference in the estimate is to multiply the CpG mutation density by a mutation rate that is approximately ten times the non-CpG rate as previously described.[9,25] In this case we calculate an average CpG mutation density for the Ya5 subfamily (241 mutations/11088 CpG bases) or 2.17%,

and (275 mutations/11,224 CpG bases) 2.45% for the Yb8 subfamily. Using a neutral rate of evolution for CpG based sequences of 1.5%/million years yields estimates of 1.44 and 1.63 million years old for the Ya5 and Yb8 Alu subfamilies, respectively. Both estimates are consistent with the initiation of the expansion of the Ya5 and Yb8 Alu subfamilies that is roughly coincident with the divergence of humans and African apes.

Inspection of the nucleotide sequences flanking each Ya5 and Yb8 Alu family member shows that most of the elements are flanked by short perfect direct repeats. The direct repeats range in size from 3-23 nucleotides. The observed direct repeats are fairly typical of recently integrated Alu family members.[7,9] The appearance of truncations within a number of these elements probably occurred as a result of incomplete reverse transcription or improper integration into the genome rather than by post-integration instability. All of the Ya5 and Yb8 Alu family members analyzed have oligo(dA)-rich tails that range in length from six nucleotides to over 60 nucleotides in length. It is also interesting to note that the 3′ oligo(dA)-rich tails of many of the elements have accumulated random mutations beginning the process of the formation of simple sequence repeats of varied sequence complexity. The oligo(dA)-rich tails and middle A-rich regions of Alu elements have previously been shown to serve as nuclei for the genesis of simple sequence repeats.[28]

Alu Y to Yb8 sequence evolution

In our query of the human genome, we identified 88 Alu elements containing one to seven of the eight Yb8 diagnostic nucleotides. These 88 "mosaic" elements were subdivided into Yb1, Yb2, Yb4, Yb5, Yb6 and Yb7 depending on the number of diagnostic changes present (Figure 1(a)). To facilitate identification of the individual elements with different diagnostic mutation combinations, the mosaic elements were numbered consecutively in order of abundance (Yb1.1, Yb1.2, etc., see Figure 1(a)). No evident sequential order of accumulation of the Yb8 diagnostic mutations can be easily discerned. Interpretation becomes complicated due to the fact that four out the eight diagnostic mutations are CpG changes (positions 1, 2, 4 and 6 Figure 1(a)). The Alu Y has three CpG sites (positions 1, 2 and 6) that become TpG in Yb8, and Alu Yb8 has one (position 4). CpG dinucleotides mutate at a rate that is about 9.2 times faster than non-CpG,[9,25] as a result of the deamination of 5-methylcytosine.[26] Therefore, it is difficult to know if the presence of a TpG diagnostic mutation is due to a change in the Alu source gene or in the particular individual Alu element being evaluated. Because CpG dinucleotides represent hot spots for mutation, a high proportion of CpG positions in the Y subfamily might have mutated to TpG. This makes discrimination between source gene changes and parallel forward mutations occurring in multiple Y elements at these loci difficult. Therefore, we have eliminated these sites (positions 1, 2 and 6) from our analysis (Figure 1(b)). Position 4 represents a different situation. Because the TpG to CpG mutation occurs at the normal evolutionary rate, it was not eliminated from the analysis. However, some variations may be observed where individual copies might have mutated the position back to a TpG that need to be taken into consideration. Now, a sequential evolution of the appear-
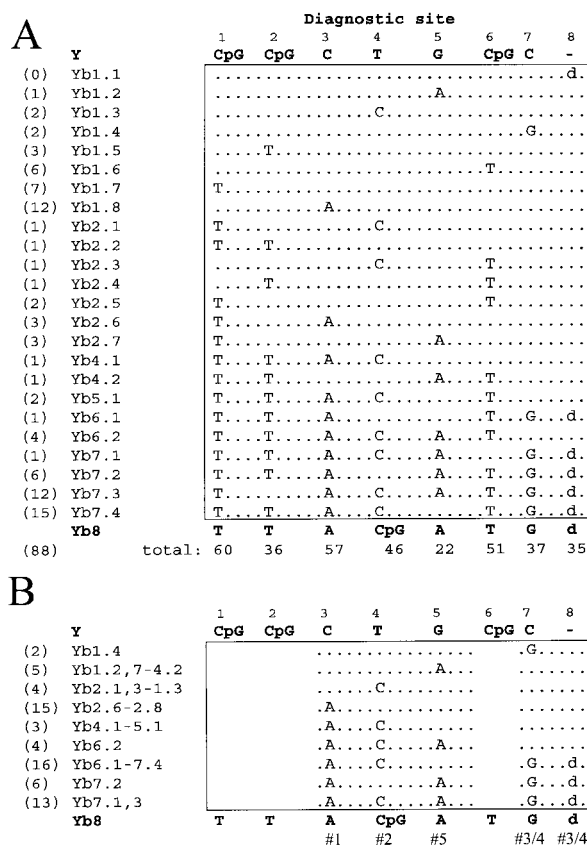
```
A                          Diagnostic site
                  1    2    3    4    5    6    7    8
           Y     CpG  CpG   C    T    G   CpG   C    -
 (0)  Yb1.1   .........................................d.
 (1)  Yb1.2   ..........................A...............
 (2)  Yb1.3   ..............C..........................
 (2)  Yb1.4   ...............................G.....
 (3)  Yb1.5   .....T...................................
 (6)  Yb1.6   ...............................T.......
 (7)  Yb1.7   T........................................
(12)  Yb1.8   ..........A..............................
 (1)  Yb2.1   T...............C........................
 (1)  Yb2.2   T....T...................................
 (1)  Yb2.3   ..............C.........T................
 (1)  Yb2.4   .....T.................T.................
 (2)  Yb2.5   T.....................T.................
 (3)  Yb2.6   T..........A............................
 (3)  Yb2.7   T..................A.....................
 (1)  Yb4.1   T....T....A....C........................
 (1)  Yb4.2   T....T...........A....T.................
 (2)  Yb5.1   T....T....A....C........T................
 (1)  Yb6.1   T....T....A..............T...G...d.
 (4)  Yb6.2   T....T.....A....C.....A....T........
 (1)  Yb7.1   T....T....A....C....A......G...d.
 (6)  Yb7.2   T....T....A....C.....A...T...G...d.
(12)  Yb7.3   T..........A....C.....A....T...G...d.
(15)  Yb7.4   T....T....A....C...........T...G...d.
      Yb8      T    T    A   CpG   A    T    G    d
(88)  total:  60   36   57   46   22   51   37   35

B                 1    2    3    4    5    6    7    8
           Y     CpG  CpG   C    T    G   CpG   C    -
 (2)  Yb1.4           ................      .G.....
 (5)  Yb1.2,7-4.2     .............A...      .......
 (4)  Yb2.1,3-1.3     ......C.........      .......
(15)  Yb2.6-2.8       .A..............      .......
 (3)  Yb4.1-5.1       .A....C.........      .......
 (4)  Yb6.2           .A....C.....A...      .......
(16)  Yb6.1-7.4       .A....C.........      .G...d.
 (6)  Yb7.2           .A..........A...      .G...d.
(13)  Yb7.1,3         .A....C.....A...      .G...d.
      Yb8      T    T    A   CpG   A    T    G    d
                           #1   #2   #5        #3/4 #3/4
```

Figure 1. Evolution of the diagnostic nucleotide positions from Y to Yb8 Alu elements. (a) Alignment of the eight Alu Yb8 diagnostic nucleotides and the different Yb1, 2, 3, 4, etc. elements found in the databases. The eight diagnostic nucleotides are indicated in bold at the top for Alu Y, and for Alu Yb8 at the bottom. At position 8, − or d represents the absence or presence of the seven nucleotide duplication, respectively. For easy reference, individual elements containing different combinations of the diagnostic mutations were numbered consecutively in order of abundance (Yb1.1, Yb1.2 , etc.). The total number of elements found for each subgroup is indicated on the left in parenthesis. Note that no Yb1.1 was found (0). The total number of the Yb8 individual diagnostic sites found in all the intermediate elements is indicated at the bottom. (b) Alignment of the same elements after eliminating the diagnostic sites in Alu Y elements involving CpG to T changes. Commas separate elements within the same Yb group and dashes between different groups, i.e. Yb1.2,7-4.2 represents Yb1.2, Yb1.7 and Yb4.2. The suggested evolutionary order of the occurrence of the changes at the diagnostic sites are indicated at the bottom (#1, #2 . . .).

ance of the diagnostic sites can be obtained, starting with position 3, then 4, 7 and/or 8, and finally position 5 (Figure 1(b)). The mutation at position 3 appears to have occurred first, being the most common single nucleotide change with 15 Yb8 mosaic elements. The other Alu Yb8 mosaic elements with only one diagnostic nucleotide change occur in lower frequencies and may be explained by parallel mutations, post-transcriptional selection,[8] or by a forward gene conversion event. The order in which the mutation at positions 7 and 8 (the seven nucleotide duplication) occurred cannot be resolved with these data. Four of the elements (Yb6.2 in Figure 1(b)) do not fit the proposed sequential evolutionary pattern. In this case multiple recombination events would be required to obtain this outcome or some selection occurring at the retroposition process, both highly unlikely. Alternatively, position 5 may be explained by gene conversion events or parallel mutations. The possibility of gene conversion between Alu repeats has been suggested previously.[29] In addition, limited amounts of gene conversion between Yb8 Alu elements[21,30] and extensive levels of short gene conversions in the Ya5 subfamily[18] have been previously reported.

Phylogenetic origin

In order to determine the approximate time of origin of each Alu subfamily member (Ya5 and Yb8) in the primate lineage, we amplified a series of human and non-human primate DNA samples using the polymerase chain reaction (PCR) and the oligonucleotide primers shown in Tables 1 and 2. In this assay, genomes that are homozygous for the presence of an Alu element amplify a PCR product about 400 bases in length. Genomes that do not contain the Alu element at a particular chromosomal location amplify a 100 bp fragment, while heterozygous genomes amplify both fragments. Using this approach we investigated the phylogenetic origin of each Alu element. All 231 Ya5 Alu family members were subjected to this analysis and only one element (Ya5NBC42) was present in the orthologous locus from the common chimpanzee genome. For the Yb8 subfamily, 244 elements were assayed with one (Yb8NBC253) being present in the common chimpanzee genome. This suggests that almost all of these Alu elements dispersed within the human genome sometime after the human and African ape divergence and that less than 0.42 % (2/475) of the Ya5 and Yb8 Alu subfamily members in the human genome also reside in non-human primate genomes.

Human genomic diversity

In order to determine the human genomic variation associated with each of the Ya5 and Yb8 Alu family members, each element was subjected to PCR amplification (outlined above) on a panel of human DNA samples. The panel was composed of 20 individuals of European origin, 20 African Americans, 20 Greenland Natives or Asians and 20 Egyptians for a total of 80 individuals (160 chromosomes). Using this approach 134 Alu Ya5 (Table 1) and 160 Yb8 (Table 2) subfamily members were monomorphic for the presence of the Alu element, suggesting that these elements integrated in the genome prior to the radiation of extant humans. A total of 28 Ya5 and Yb8 Alu family members appeared heterozygous in all of the individuals that were analyzed, suggesting that they had integrated into previously undefined repeated regions within the human genome as reported previously.[31] In the PCR-based assay these elements generate a pre-integration site size product from the duplicate copies of the pre-integration site located throughout the genome along with an Alu filled site from the one pre-integration site sequence that contains the new Alu insertion. These elements were not subjected to any further analysis. An additional six elements were located in other repetitive regions of the genome that were identified computationally and discarded from further analysis. The remaining elements were polymorphic for the presence of an Alu repeat within the genomes of the test panel individuals (Tables 3 and 4). Loci that were polymorphic for the presence/absence of individual Alu insertions were subsequently classified as high, low or intermediate frequency insertion polymorphisms (defined in Tables 1 and 2). The unbiased heterozygosity values (corrected for small sample sizes) for these polymorphic Alu insertions were variable, and approached the theoretical maximum of 50 % in several cases. This suggests that many of these Alu insertion polymorphisms will make excellent markers for the study of human population genetics. Approximately 25 % (58/231) of the randomly identified Ya5 and 20 % (48/244) of the Yb8 Alu family members are polymorphic for insertion presence/absence within the human genome. These results are in good agreement with previous estimates of the percentages of insertion polymorphisms within these two Alu subfamilies.[21]

The Alu inserts that have been in the genome longest are more likely to approach fixation. Therefore, we might expect to find different levels of sequence divergence for the Alu elements from each insertion frequency class. Using this approach the average number of non-CpG/CpG-based mutations for the Ya5 Alu family was 1.62/1.06, 2.83/0.67, 2.16/0.66 and 2.53/1.0 for the fixed present, high frequency, intermediate frequency and low frequency Alu insertion polymorphisms, respectively. In the case of the Yb8 subfamily the average number of non-CpG/CpG mutations was 1.86/1.16, 5.0/0.6, 2.2/0.66 and 1.7/1.2 for the fixed present, high frequency, intermediate frequency and low frequency Alu insertion polymorphisms, respectively. In all cases the standard deviations for each average were as large or larger than the average number of mutations reflecting the heterogeneity in the dataset. No detectable

difference in the mutation density within each frequency class of Alu insertions was observed. Therefore, our data suggest that any sequence differences between the polymorphic elements and those with fixed presence may be obscured because of the small number of total mutations and sequencing errors (see Discussion).

Discussion

Alu elements account for more than 10% of the mass of the human genome. The majority of Alu elements integrated into the genome early in primate evolution. Only a small number of elements (a few thousand) have amplified in the human genome after the divergence of humans and African apes. Here, we report an investigation of the dispersion and insertion polymorphism of the two largest subfamilies of recently integrated Alu repeats within the human genome. Our copy number estimates of 2640 Ya5 and 1852 Yb8 Alu elements within the draft sequence of the human genome are in fairly good agreement with previous estimates of the sizes of these Alu subfamilies although they both exceed the previously published figures.[21]

Using the mutation density and a neutral mutation rate we were able to estimate the ages of each subfamily as 5.32 million years (myr) old for Ya5 and 5.30 myr old for Yb8 using non-CpG-based estimates and 1.44 myr (Ya5) and 1.71 myr (Yb8) using the CpG mutation density. Each of these reported average ages based upon non-CpG mutation density is substantially higher than those reported previously of about 1 myr and 2.7 to 4.1 myr for the Ya5 and Yb8 subfamilies, while the estimates based upon CpG mutation density compare favorably to those previously reported.[21,32] If we assume a linear amplification of these Alu subfamilies in the human genome, the oldest elements would be no greater than 10.64 myr old for Ya5 and 10.6 myr old for Yb8 using non-CpG mutation density, or 2.88 myr old for Ya5 and 3.42 myr old for Yb8 using the CpG mutation density. The non-CpG based estimates for the oldest subfamily members appears to be somewhat higher than expected for a group of repeated DNA sequences that largely amplified within the human genome after the divergence of humans and African apes which is thought to have occurred within the last 4-6 myr.[27] This discrepancy between the two estimates can be explained by considering sequencing errors as a potential factor influencing our current calculations. In the determination of the non-CpG mutations for the estimation of the Alu subfamily age, sequencing errors would be included in the count as mutations, making the estimated age higher than the actual age for the subfamily. If we assume that the sequencing errors are distributed evenly across the entire Alu sequence, then the number of sequencing errors would be higher in the non-CpG-based estimates than the CpG-based

estimates, since there are more non-CpG (242-246) than CpG (only 44-48) nucleotides in the subfamily consensus sequences. Our observation that the levels of sequence divergence from the subfamily consensus sequences do not effectively correlate with polymorphism levels in the human genome also argues that it will not be beneficial to use sequence divergence from the subfamily consensus sequences as a method for the identification of additional polymorphic members of these Alu subfamilies.

We can also compare the calculated ages of each Alu subfamily based upon CpG mutation density as a whole to the estimated percentages of Alu insertion polymorphisms and copy number to evaluate the contribution that these elements make to human genomic diversity. Here, we report estimated ages of 1.44 myr for the Ya5 subfamily and 1.71 myr for the Yb8 subfamily. The percentage of Alu insertion polymorphisms in each of the subfamilies was 25% for the Ya5 subfamily and 20% for the Yb8 subfamily. The copy numbers of the two subfamilies of Alu elements were also different with 2640 Ya5 Alu elements and 1852 Yb8 elements. When considered together these data indicate that the Ya5 Alu subfamily with both a higher copy number and more insertion polymorphisms has been more successful at amplification within the human genome. In fact, if we assume that the ages of the two subfamilies are about the same the Ya5 subfamily has been about 40% more efficient at amplification in terms of both copy number and the generation of new Alu insertion polymorphisms within the human genome. Although the sample size is presently small, this is also in good agreement with the number of previously reported Ya5 (six) and Yb8 (three) Alu repeats associated with different human diseases (reviewed in ref. 22). In addition, these data also provide compelling support for the simultaneous expansion of multiple Alu subfamilies within the human genome. The reasons for the differential amplification of the two Alu subfamilies remain unknown. However, they likely reside in the ability of each subfamily to produce RNA for retroposition or at some other point in the process of retroposition itself such as the reverse transcription step. Further experiments will be required to determine the precise molecular mechanism(s) leading to the differential expansion of these two Alu subfamilies within the human genome.

Using the non-CpG-based average ages of the Ya5 and Yb8 Alu subfamilies along with a linear amplification rate we can also estimate the number of members from each Alu subfamily that should be present within the orthologous loci of the non-human primate genomes. Using this approach the oldest Alu repeats from each subfamily would be approximately twice the average age. In other words, the Ya5 subfamily would have begun to expand 10.64 myr ago with the Yb8 subfamily having expanded about 10.6 myr ago. If we assume that humans and African apes diverged from each

Table 1. Alu Ya5 accession numbers, locations, human diversity, oligonucleotide primers and PCR parameters

Name	Accession	5′ Primer sequence (5′-3′)	3′ Primer sequence (5′-3′)	A.T.[a]	Human diversity[b]	Chr.[c] Loc.	Product size[d] Filled-empty
Ya5NBC2	M28713	CTTTAGACTACAGTTGTGTTAGCCTCTTG	CTGCACTTTCCAAATTTTCTACCAC	55	FP	22	710-384
Ya5NBC3	AL023807	ACTGCTTGAAGCTAGAACTTAAGAGACC	CTCTTGTCTGCTTCTAGACTTGTGAATAAC	60	FP	6	564-243
Ya5NBC4	AL008629	CTGATGAGAAATCTGCTGCTATTG	GCAAACCTCAAACAGGATAAACAC	60	R/R	X	483-154
Ya5NBC5	AC007363	TAGGATATTACTGTACAAAGCCGTAGATTT	GTTTTAAGCTAAGCGTTATTACAAAAGAGT	60	IF	2	476-163
Ya5NBC6	AC006344	GATTACATCCTGTGATCCTGAAACT	GAACATTTGTTCTTTTGTGACTGCT	60	FP	6	539-189
Ya5NBC8	AC006478	ATTAAGCACCAGGAAATTGCCATAC	CTAGTAAGGCTAGTCCCATAATTTGAAGTG	55	FP	7	513-196
Ya5NBC9	AC006382	CTTCCCTAGGATTTAAGTCACCATAAAGAC	TTTTCAACTTGTAACTGTAGAGGACAGGAC	60	FP	Y	415-102
Ya5NBC10	AC008725	AAAGCATAAAGAAAAGTACGCCAAC	CAATGAAGATATAGAACAGCCCCTA	60	FP	20	449-141
Ya5NBC12	AC009307.6	GGAGTCAAAGGTACTTACAACGTCT	CTCCCTGTCTTCCTAACGTAATTTT	60	FP	2	470-129
Ya5NBC13	AL031302	CTTCCTGTGTATACTTCTCTGCAC	GTCTGTGACCTGCAACACAAG	60	FP	22	604-291
Ya5NBC14	AL050342	Alu flanked by other repeats	Alu flanked by other repeats	-	-	1	-
Ya5NBC15	AC008676	CTTTCCTCAGCTTGGTTTATTCTACTG	GAAGATTCAGGTGGACAGAAAAAC	60	-	5	502-192
Ya5NBC16	AC008608	CTTAACCAAAATAGTGGACGAGGTT	CAGAAGTATTTACTACTCGAACAGTGAGC	65	HF	5	539-229
Ya5NBC17	AC007076	TATGCTCTGAGAGGTTTTCTAGATCTCTG	GAATAGGAGCATCATTCAAGTTCAG	65	FP	7	553-232
Ya5NBC18	AC008433	AGACCATCTTTAAGAGGAATACCATGT	GAAACGTAAATTTGTTAATAAAGTGGTGAC	65	HF	5	495-180
Ya5NBC19	AL109948.1	AATCACTGTTACTCATGGGGTATCT	AGACCTACGTGCCTTACTCACTGT	60	FP	1	433-116
Ya5NBC21	AC008482	AGCTGTCGTTCAAATGAGACTTTCT	AAGCTCACTCATCAATAAGAACACC	60	FP	5	512-177
Ya5NBC22	AC004519	TCTGTGTTCTTTGAATGTGTATTACTCTTA	GAATGTAAAGCTGTAACTTTCCTTTTCAAT	55	IF	7	471-156
Ya5NBC24	AJ011932	AAAATTGAGAGACGAGGAGAAGGT	CCTCATCAATACTGTAACTGTCACAAC	60	LF	21	595-286
Ya5NBC25	AC004220	GTGAAGGACATGAACAGACACTTCT	CACCAACAGTGTAAAAGTGTTCCTA	55	FP	5	538-218
Ya5NBC26	AP000311	GGGCTATTCTGATTTTCTTCTTCTC	AGAAGAGACATCACTACAGATACTACAGAC	55	FP	21	476-158
Ya5NBC27	AC003691	CTGAATACAGGTATCACTGAACAGAAC	ACAGTGTAAAGTCTAACCTACCAGAGGAT	55	IF	11	591-265
Ya5NBC28	AC005862	GGGTACATGTGCAGGTTTCTTATAC	GCTAACTGATGAGAACACACAGATACATAG	55	LF	7	474-191
Ya5NBC30	AC007159a	CTGGCACATAGTAGGTGTTCAATAA	GAGTAGTTTGAGTCTGTTTGTAGCAGAG	60	FP	?	502-191
Ya5NBC31	AL033543	GTATCTTGTGTGTTTCCTAACAAGACTGAG	CTCATTTTCACTTATCAGGTCTGTCTC	60	FP	22	523-238
Ya5NBC33	AC006288	GCAATTGCTATCCTTGAGTGTTTC	CTCCTAGTCTAGAGTTTTTCCCATTGTATC	60	FP	9	543-226
Ya5NBC34	AL031575	CACTCTGATACTTATCTCTGTGCCTGTAT	TGAGAGACATCAAACCAGAAATCC	60	FP	X	494-150
Ya5NBC35	AC004534	GAGAAGTACTCAGAGAGGACATCATTT	GTAGTCATGGAGGTAAGAAAAGAAGACAC	60	IF	7	515-179
Ya5NBC36	AC004006	ATGAAATAACTCCTAGATTCAGGCTTC	AGTTTCTTGTGTTAGTTTCCTTAAATACCT	60	R/R	7	515-200
Ya5NBC37	AC002476	GCTTGAGGTTTTCATACTACTCTTATCTTT	ACTGTATAAGCATTTTCCTCTTTATCTTTC	60	IF	X	497-184
Ya5NBC38	AC006033	GTACCCTCTAATTTACAGTCATCTCATACC	GAACTTCTCTGGCTTGAAAAATCAG	60	LF	7	487-170
Ya5NBC39	AC005533	TGGGACTTAGCTGTTTTGGTATCTA	CTAAACACAGGTTACAGCACCTCTT	60	FP	14	469-152
Ya5NBC40	AC008887	ATTGATCTCCAACTGATGCCCTA	GACAACAGACTTACCCTGCCTATACTATT	58	-	5	417-105
Ya5NBC41	AC008828	CTCTTTATGGGACTTGACAAGCA	GTTCTACATTGCCATAATAGTGTAGGG	55	FP	5	441-128
Ya5NBC42	AL078621	AGTAAGTCCCTCCCCATATGCT	GGTCTTTCTAACCCAAAGGTCAC	55	FP	22	486-185
Ya5NBC43	AL096867.7	CCTTTCCTTACTAGACAGTGACAACAT	CTTTTAGCCATCTTCTTGGTGTTTG	55	FP	6	539-218
Ya5NBC44	AL096840	CATTCTCTCCTTTGGTCCTCTTTAT	GTGAGTTTGGGGATATGGTGAG	55	FP	1	525-202
Ya5NBC45	AL049868	TAGGGTAAGGAATATGTGCTGCTTTAG	GTCTCTGAACGACTATGTGAGCAG	60	IF	20	591-265
Ya5NBC46	AC009466	GATGTGTGAATACTGTGTAGATTCCAG	GTAAAGCTTTTGTAGTGCCTAGCTCTTAGT	55	FP	11	405-94
Ya5NBC47	AC007227	CTCAAGATTGGCCTATAGTCGTTAT	AGACACAGGTATCCAGTGAAAGAGT	55	FP	1	526-200
Ya5NBC48	AC002290	ACTGTTAAGATAGTGAATTTTACTGCTCCA	AACTCACACGTGATACAAAATACTCACAGA	55	FP	11	481-176
Ya5NBC50	AL096829	AGCCTGGCCCATGATAAAAA	GTTTCATAGATAAGACACTGGCATGTTACT	-	-	1	465-155
Ya5NBC51	AC008249	ATATTCCAGAAGTTTCCTTACATCTAGTGC	AAAGCTTTAAGTCTCCACCATCTCT	60	IF	3	437-140
Ya5NBC52	AC009094	AAGTTGAAGTCCAAAGTCTCATCTG	CTTCTTCCCCACTTAAATATCAAGG	53	-	16	688-371
Ya5NBC54	AL024507	GTTTATGTCAGTAGGAGTTTTCTCGTGTAG	TCATTGTATCATCTGCTGTACCTGT	60	LF	6	433-130

Name	Accession	Primer 1	Primer 2				Range
Ya5NBC56	AL109767	TCATTGTATCATCTGCTGTACCTGT	AGTCCAACATAGATGTAATTGGGAGTTCAGG	60	FP	14	469-148
Ya5NBC57	AC009107	GACGTAAAGAGATGTTGTTAAGTGAAAAAT	ACTGTAGGAGGTAAATGGAAAGTCAACAGA	60	IF	16	444-126
Ya5NBC58	AC008376	TGCCTCTTTAACCAATTCTCTTATTTTCA	TATTTGGCTGGATTTGAGTTATCTCTTAGG	60	FP	5	481-141
Ya5NBC61	AC009594	TGAAATAATCCAGTTGGGGAAG	GTATATCTCTACCGAGACTCAGTTTTTAGC	55	IF	4	493-180
Ya5NBC66	AC006210	ATGGTAATTTCCCTCATTTGTCA	GTAATGTCCTCCATTGTTCATTTG	61	FP	X	448-115
Ya5NBC67	AC006005	CACACACCCCGTATTTTCT	TGCATTCTCTTGGAGTTTG	58	FP	7	424-131
Ya5NBC69	AC004053	GGGATCAGTTACAGTGCTTC	ATGGAACGCAACTTAGAACT	50	FP	4	359-42
Ya5NBC70	AC004454b	ATCAACGTGGGACATAACCA	TTCAGAAGGCACATTAGTGCT	60	FP	4	391-116
Ya5NBC72	HS234H5	CCTTGCTGCATAAAAACCTA	TATGACTAATGTGGGGCTTT	52	FP	6	416-102
Ya5NBC73	AC004454a	GACTATTAATACGAATCCAAAGTACACG	TTTAACTTGGTCCTACCTGTGTGTC	51	FP	4	465-129
Ya5NBC76	M96868	ACTCTTTAGTTTGTAAGATGGCAAG	GGTGGAGGGAAGTAGCAGAA	60	-	2	735-417
Ya5NBC77	AL008629	TTTTGCCGTTAGTTTTCAGAAG	GGGCAAACCTCAAACAGGAT	60	-	X	408-83
Ya5NBC78	AC006155	AACTCCAACAGCCACATCCT	TGGTGGGTCAGTATTGAGTGA	60	FP	7	382-66
Ya5NBC80	HS960017	CTCTCCTGTGTCCATACTTCTT	CTGGCATGGAGATTTCTTAC	60	-	X	368-47
Ya5NBC81	HU95742	GTGGCAGTGGAGATAGAAAA	ACACTCAATCCATCACCTTT	60	-	16	352-42
Ya5NBC82	AC005217	AGTGCTGGGAAGACCAGT	GCTCGTGCTCTTGACAAA	60	FP	5	348-37
Ya5NBC83	AL022101	AGGGAAGTCAGGCAGAAC	CCTTCCCTAGGAGCACAT	55	R/R	1	384-106
Ya5NBC87	AL109830	AGTAAATACTGTAAGAGATGTGGAGAGCAC	CACAGCATTAGAGAGAGTTGATGATAG	55	FP	20	434-201
Ya5NBC89	AC009807	ATCTTCCCGGCCATAAACCTC	GAGGCCCAAATTTGCTTACA	55	R/R	11	516-195
Ya5NBC91	AL034378.2|	GGTCATGGTTCTTTGCTATTCATTC	GTATTGTAACCCATAGAGCAACCAT	55	-	1	531-214
Ya5NBC93	AC010086.1	CCTCTGCCATATGGTATGTTTAGAA	TAGGATGAGGTCAAAAGTGAGAAAC	60	FP	Y	610-294
Ya5NBC94	AC008788.1	ATTGTCTTTCCTGTGCTACTCTCAT	CACTTTAGTGGATGCTTATCTTTG	57	FP	5	531-201
Ya5NBC95	AC009962	ATTATAGTAATAAATCCTGGAATTGGACCT	GAAATGAATCTATGGGATATGTTTCT	60	IF	2	489-148
Ya5NBC96	AC004547	TAGATGAGATAGAGCCATCAAACACTC	GTATTGTAACCCATAGAGCAACCAT	60	IF	7	509-169
Ya5NBC97	AC004453	GCTCTTTTCTTGTTTTCTGGAAGTG	TGTGAGTGTAAGAACACGTGTAAGAG	55	HF	7	442-147
Ya5NBC98	AL049591	TATAGCTAGTAAATGGTAGAGCCAGGA	CTGTCTAAGATAGTAGTTGGACCTACTATG	62	FP	X	504-209
Ya5NBC99	AL031312	TATACACACACACAGAGAATGACTG	CCTGACTCGAAAGTACTGTTTTCTAAG	55	-	X	515-198
Ya5NBC100	AL035683	TGAACAGTTCTCTTAGGTGGTTAGTAG	TAATATACAGTTGTGCTCACTAGCATACCC	50	FP	20	477-153
Ya5NBC101	AC006030	CTCACTGACACTTTTGGTCAGACT	ATTTACTGAGCACAATGCCTCATAC	55	IF	2	519-204
Ya5NBC102	AF118569	TCCCATTTCTTCTAGACCTGCTG	CCCATAACAGGTCTTCATATTTCC	55	FP	17	483-194
Ya5NBC103	AL034408	ACTCTCTCCTACATCACTGACTTCTC	GTAAGCTTGAGTTCAGAGGACAGATA	58	FP	X	556-237
Ya5NBC104	AC07065	GGGCATAGCTGTAGATATAAGCACTACAA	AGAAGAATAGAGGACTATGTCTGTGTCTC	58	FP	12	508-188
Ya5NBC105	AC006040	GTATACATTCTGCAACCAGTGGAG	AGAGGTAGAGAGCTTGCATTTCAG	62	R/R	Y	598-281
Ya5NBC106	AC005532	CTTACAGTTACGGAGGCTAGAAAGTC	GTTATCATGGGAAGGGAACTGT	60	R/R	7	509-207
Ya5NBC107	AC004884	GTAATGAATAGCCGTGCAACTGTC	CACCCAGCCATTTCCTAGTTAT	60	FP	7	556-236
Ya5NBC108	AC007092	CATATGAGTGCCTGACTTTTACTACTTCTC	CTAAATACAGGATGAAAGGACTGGTAG	60	FP	2	567-215
Ya5NBC109	AC005745	GTGCCTGGTACTCTAGAAATAAAACTCTCT	AGAATGAAACTCCGGCTCAAA	58	IF	22	561-251
Ya5NBC110	AC004761	GAGTCTTTGTTCTGTTAACTTAGTGGTGAG	CTAGAAGGTCACACATATGTCAAGGAT	60	FP	5	558-170
Ya5NBC112	AC008032	GGTTATTAGTTTTGGGGTGGTAGTC	GGGATACCCAATTCAGTTGTACTAGA	60	FP	3	396-93
Ya5NBC114	AC007782	GAGGATTTTAGCAGATAGTATTGTGTTACAG	CAAAGTCTTATCAAATACAGCACACTG	60	-	12	524-223
Ya5NBC115	AC009316	AGAACAAACTGCACATGCGAGTATCT	ACCTTCAAATTTCTCCTTTGAGGAC	60	FP	2	574-240
Ya5NBC116	AC006344.2	GATCCTGAAACTATTTAAAATCAAGAAGAC	TCTAACCCATATGTAGAGTTAATCTCTTTGAC	60	FP	7	575-237
Ya5NBC117	AJ010770	GGGAAGGAGAAAGGAGAAACCATCTAGT	CTCTCCAGCTATAACCCCAACTACT	60	FP	7	515-188
Ya5NBC118	AC005913	AATACGTGTGTCTGTGTATATGTTT	TGCATAACCTTCCCAGAGATAATG	60	FP	X	533-235
Ya5NBC119	AC006002	TGTTAATAAACAAGAACACTACTCCAAGG	CTTTTGTTATATATACTGAGGAGAAAATGG	60	FP	7	482-167
Ya5NBC120	AC005863	GGACCACATGACTGAGTGTAAAGT	GAGGTGGCCTCTTAACCATAATTC	55	IF	17	518-199
Ya5NBC121	AJ011932	AGGGGGAAAAACATCAAAAACTC	CCTCATCAATACTGTAACTGTCACAAC	60	FP	21	510-202
Ya5NBC122	AC005747	CCATTCATTCTATTTGGGGAGTTAG	GACTAAACCAGGATGTGAGCTTTT	53	FP	17	527-217
Ya5NBC123	AC005739	ATCAAGTTGACACTCAGTATTCACCAC	CTAGTCTGCAGAACTGTGAGAAATGTA	55	IF	5	490-180
Ya5NBC124	AL022310	CTAGACAGTGCAACAGTTCCTAAATACAG	CATAATGAAAACTCCATGTGCTAC	60	FP	1	457-131

Table 1. (*continued*)

Name	Accession	5' Primer sequence (5'-3')	3' Primer sequence (5'-3')	A.T.[a]	Human diversity[b]	Chr.[c] Loc.	Product size[d] Filled-empty
Ya5NBC125	AC004206	AGTATTTTGCACTTCTCTCAAGGGTGTC	CTGGTCTTATGTTTCATCTGGATTC	60	FP	6	507-223
Ya5NBC126	AC005144	GTCTGCTGAATGATTAAACCAACAC	GTGCCATTTCTACTACTGAAACCTAAC	60	FP	17	480-171
Ya5NBC128	AC004808	GGGTGGGACAAAGAAATACTCA	GCTTATGGCTTGCAGTTTCACT	55	-	7	648-293
Ya5NBC129	AL008635	TACATGGAGTTAGAGCCCGTTC	ACAAGTGGCTGTCACYCAACAC	60	FP	22	486-180
Ya5NBC130	AC004629	GTTGTGTCCACTCTTGACTAGTATGA	GACAGTTTACTGACTACACAGGATTTCAG	60	FP	5	602-287
Ya5NBC131	AF002996	CCCAAGATCTAGGTGATGGACAC	GCACTTGAGATAACCTAGTTAGAATGC	60	IF	X	495-174
Ya5NBC132	U91328	CTCGTGATTCACAGAAGTGTTGTAAG	CGGGGTTCATCCTTAATACATACAT	60	IF	6	458-228
Ya5NBC133	AC000355	TGTTATCATTACCAAATACAGCCACTTTAG	TCTTTGGCTATAAGGATATGAAAACTTAAC	60	FP	7	692-374
Ya5NBC135	U01102	ATTAAGCTCATGGTAACCAGCAC	GACTCTCCTCTCTGGATTAGAAACAG	60	LF	11	436-117
Ya5NBC136	AC008124	CAGCAACAATCAAAGTTTATAATGC	GGAAAATTGAATGATGGCAAA	60	FP	12	749-439
Ya5NBC137	AC005002	GTTGCTGTTTTCTGCTGCAC	GCATAAGAGACCAATCCTGGAG	55	R/R	?	521-197
Ya5NBC139	AL031650	TGAAAGCTCTTAAGGTCTTCTCTCT	TAAGTAGACCAGAAACAGGGAACAG	60	FP	20	851-634
Ya5NBC140	AC007877	GCAGCCCCAAGTGTTAAATTACTAT	GGTTGTGGTAATGTCATCATAAACG	60	FP	2	471-135
Ya5NBC141	AL096769	CTGAGAAACCAGCAAAGTAACTGAC	CATGGAACCCATATACAGACTACAAA	60	R/R	20	480-139
Ya5NBC142	AC007392	ACATTCTAGGACACCTGTCAGTCAT	GGTCAATAGCATGGGAAAGAATATC	60	FP	2	663-321
Ya5NBC143	AC006374	GCAATGCACATAAGATATGCTC	CTTTTCCCTACCATGGTGTCTTT	60	FP	7	572-251
Ya5NBC145	AL035667	TGCATCCTTCTTCTGCTGTTC	AATTGGGTTCACTAGACAAAGG	60	FP	20	500-276
Ya5NBC146	AL022329	CTGTCCCTTCTCAGGCTCATT	CTAGCATGTTGTCACCTCTCAACC	53	R/R	22	604-131
Ya5NBC147	AC007656	TAGCTGGGGAGGTAGATAATAAAC	AAATATCACCTTATCAGTGGGACCT	60	LF	12	493-155
Ya5NBC148	AL031659	ACAAGATGACAGATCTAAACCCAAC	AAGGTGTTGTCAGACTAATCTATCG	60	IF	20	505-193
Ya5NBC149	AL033525	GTGTTACTGTGGCCAACTATCTCAT	ACTTATATGAGCGGGGTACAGTTCT	60	FP	1	466-155
Ya5NBC150	AF135028	AAATGGAGACACAGAGGTGTAAAGA	CCCAAACTGCATATTTAAAGGGTAG	60	IF	19	491-169
Ya5NBC152	AC004953	Alu flanked by other repeats		-	-	7	-
Ya5NBC153	AC005820	CCAATCTGGGAATTATGACAAGTAG	CTTCAGACTTCTGCTTGATTTCTTC	60	FP	Y	496-186
Ya5NBC154	AC006371(B)	AAACACCCTAGATGCTGGGTAA	AGATGAGTGAGCCTCAGAACAAAG	60	IF	Y	501-197
Ya5NBC155	AC006565	TGTCAATATCAGACAGATCCATGAG	ACTTCCAACTATGTGGTCAGTTTTG	60	LF	Y	505-182
Ya5NBC156	AC002531	TGTGGTAAGTGTAGTTTCAAAAGAGTTT	TAATCTCTGGACTGGAAACATAAAA	55	FP	Y	480-148
Ya5NBC157	AC005281	CATACGTTAAATCACTCGGTACTCA	TCAGAAAGTATACAGGTGATGTGC	60	HF	17	516-207
Ya5NBC158	AC005019	TATCTCCCCTACCAAATTTCTTTC	GGATGGATTAGAAAGGATGGATTAG	60	FP	7	500-172
Ya5NBC160	AC005245	CTCAGCTGTGCCTGATACTCTATAA	GCCTACTGGATAAGTCACACATTT	55	IF	17	551-234
Ya5NBC161	AL031978	CCTGTCTAAACTCCAGAATGAAGAA	GCAGTAGAAAGATCACAGGCTCTAA	60	FP	6	491-199
Ya5NBC162	AC003957	ATGAGCAAGTCTACATATTCCTCCA	CTTGTTGCTGTCAAGGGTCTAATA	60	FP	17	481-167
Ya5NBC163	AC004057	CAAACCAAGAGTTCTTATCACCAGT	TAGTAAGAGGTTTCCAAGTACACG	60	FP	4	624-316
Ya5NBC164	AF042090	CTGCTGACTTTGAACTTAAACTGC	GATGGAAGATGTCTTAGGGTTCTCT	60	FP	21	503-190
Ya5NBC166	AC004040	CCCTTGGCTCTATAGATAAAGTTGG	ACTGCACCAAAACTAGAGAGGAAA	60	FP	1	532-210
Ya5NBC167	AC003980	AGCCCACAGCTAACGTTATACTAGA	GTGGGGTCTTTAAGGTTTCAATAG	60	FP	7	515-239
Ya5NBC168	Z97876	AGTGCTAACCAGAGATGTGTGTGAC	TTAGTGGAATGTTCCAGGACTGTAT	45	-	1	492-164
Ya5NBC169	AC002456	TATATAATCCCACAGTAAGCCCTCA	ATAGTTGTATACCAAGCCAACGACA	60	IF	7	493-184
Ya5NBC170	Z94722	GCAAGACCTGTCATACGGTGCTTAAAT	GGACTGGGCTCTTCAGGAC	60	FP	X	521-195
Ya5NBC171	AL035688	TCTAGAATTACAAGTGCAAGCCATC	CTTCTCATCCCTGCTAACATAACAT	55	LF	6	451-130
Ya5NBC172	AC006371	CCAAACGTAAGATTGAGTGG	AGTGGTGTTCTCGGTATTTC	55	LF	Y	473-155
Ya5NBC173	AC003977	ACACACACGAATGCAGGATAAT	TGCTCACAGTCCTTAGACTTTACAA	53	-	16	508-107
Ya5NBC174	AC006462	TCACTCTTTGTCTTGCTGACTACAG	GCTATAGCTTCTATTTACGGGGAAT	55	IF	Y	526-206
Ya5NBC175	AC000396	CCAGTGTCATACGGTGCTTAAATC	GGACTGGGCTCTTCAGGAC	55	FP	9	483-148
Ya5NBC176	Z74739	GGGGGAGTATGGTTTGATATACAG	CCCTCATGGGAGGTGTTATTT	55	FP	13	666-298
Ya5NBC177	AC000111	TTTTCCAAGCGGTCCCTTAT	TCATGGAAAGCTTGTTTGGT	53	R/R	7	617-300
Ya5NBC178	AC004900	AGAGCTGGACTCTGATGTTAGAC	GAGGCCATGATGGAGGAATACTAGAC	60	FP	14	583-260
Ya5NBC179	AC006373	GCAGAAGCTTGCAATAACCTCT	GCTGAACACCTAAACACTGCTGAGAC	60	FP	7	797-490
Ya5NBC180	AL109618	CTTGAAGATCGCCATGAGTAGA	GGCATTCTTTGGACTTGTCTC	55	FP	20	525-211
Ya5NBC181	AC008041	GTTACAGTGCCTACTTCTCGGTTCTC	AGCCTTCCATCCTCATAGACC	58	FP	3	450-205

Locus	Accession	Primer 1	Primer 2	Temp	Code	Chr	Size
Ya5NBC182	AC006365	GAAGGACTATGTAGTTGCAGAAGC	AACCCAGTGGAAACAGAAGATG	60	IF	7	563-287
Ya5NBC183	AC006365b	GGACAGGTAGAGACGATTTCCTAGA	CAAGGGACTCATGTACTCTGTGAAC	60	FP	7	722-410
Ya5NBC184	AC000047	CTTGGATAGAGCTGGGAGGTCATTA	ACCCAAGCAGTTTATACTGTACCC	60	LF	9	522-205
Ya5NBC185	AC006552	GAGTTTATTTGCCGTAGGTAGCTC	GGTAGGGCTAAATGGAAAACA	61	FP	4	513-202
Ya5NBC186	AL035445	CATCTTCTGAACCCATAGGGAAAAT	GCCAATTGCCTGGTATGTTTTA	55	FP	6	649-381
Ya5NBC188	AC004970	GACAAGGACACAGATGTTAGGAATC	ATCTCTTGCAGTTGAATGTCTAAGC	55	FP	7	476-156
Ya5NBC191	AC007191	TGACGGGTGAGATGTATATAGAAGC	ACTCTTCTCATCTGTGTCAATTTGG	60	FP	19	645-330
Ya5NBC192	AC005678	CACTCAGATAAAGATGTGGACTTCA	GCTTTAGAGAGTCTGACTTTGCTTC	60	FP	6	536-238
Ya5NBC193	AC005065	GTTCTTTTTCTCTAAATGCCTCCTC	CCACATTTTCTGGAACCACTTTAC	60	FP	7	525-206
Ya5NBC194	AC004866	TATTCTTATGCCGTTATGTCCTCAG	CCATGGAATACTACTCAGCTATGAA	55	LF	7	486-169
Ya5NBC196	AL031785	Alu flanked by other repeats		–	–	6	–
Ya5NBC197	AL031785b	CAGAAGTAAGATTCGTCGGATCGTAT	CTCAATGAGATATCACCTCACACAT	60	LF	6	461-204
Ya5NBC198	AC004055	GCATAACTCCTAACCCATAATTTCC	GATCTAACACAACCAACTCCATCTT	60	FP	4	530-230
Ya5NBC199	AC005293	CTACCATCAATAACTTGGACACAGA	ATTACAGAGAGCCTGCCATGAT	60	FP	12	500-200
Ya5NBC200	AC005161	GTTTAATGGAAACCACCAAACGTAG	AGGCTGCTAGTTTCAGAAGGATAAT	50	R/R	7	500-174
Ya5NBC201	AC004745	CGCCACTTTCCCCAGTTA	CACCTCCCTAAAAAGCAGGA	50	IF	7	499-188
Ya5NBC202	AC004603	ACGCTCCAAAGTCCTCACCT	TGGAAGCTGGTTCTTCAGTG	60	FP	19	487-154
Ya5NBC203	AC004593	CAGCCTGTAGAGCTGGAAAAG	ATACAACAGTTCTGGAGGTCTGAAG	55	LF	7	445-128
Ya5NBC204	AC002385	AAGCAAATCAGTCCTACCATGA	TATTTTGGAGAGTTGTAGGCAGGA	55	LF	7	519-186
Ya5NBC210	AC004848	GAGGGGGTAGGGATAGCATT	GTGTTAATATTGTCCCACATGTAA	62	IF	7	750-424
Ya5NBC212	AC002074	CATTTGGCGCAAGTGGTATT	ATCCCAAAGAAACCCACGA	60	HF	7	502-190
Ya5NBC213	AL078463	AGGAGAGTGGGAATGGGGTGA	GATTCTCATGTACCCATCATGC	60	R/R	1	397-91
Ya5NBC214	AC004948	TGTTGTTGCAAAGGACAGGA	ACGTCCACATTCCCATGTTT	55	LF	7	500-170
Ya5NBC215	AL096710	GCCAATCTAAACGAATAATCA	AGGCAGAATGATGTTGTTGG	55	FP	6	780-467
Ya5NBC216	AC007245	GATGTGACCCTGGCTTGTAAA	CAGAGTCCCTGTGCAAAATG	55	IF	7	456-141
Ya5NBC217	AC007298	TCCAAACCTTTTTGCTCTGC	GTATTTTGCCCCTGCCCTA	60	–	12	623-308
Ya5NBC218	AC006989	AGCCCAACATCTGGTTTTGT	TCCAGTCTCGTGTAAAATAGCTTG	55	FP	Y	445-109
Ya5NBC219	AC006989b	CAGTTTTCCATATACATGTGGGTTC	AAACCTGGAGGGGCATTCTTT	58	IF	Y	445-129
Ya5NBC221	AC004019		TAGTTGTTAAGAGGCCCATTTTCTAC	60	IF	22	640-313
Ya5NBC223	AC005006	GTTCTCTGTAAAATGGACCAATCAG	CATAGACCCCCAGTGAGTTAC	60	LF	22	455-214
Ya5NBC224	BK407F11	ACATGCTTTTCCCATTATGTGTG	CCAAGTGGCAGTAATAGACTCTGTC	55	FP	22	502-195
Ya5NBC225	AC002470	Alu flanked by other repeats		–	–	22	–
Ya5NBC226	DJ323M22	CCTCCACGGACTCCTAATTACA	GTGGCCCTGAGAAGGAATTT	55	FP	22	421-130
Ya5NBC228	AC004832	ACTGCATGCCAGCCTCA	GCTAGTTACAATGAAAATGTGCTGT	55	FP	22	842-529
Ya5NBC229	AL096873	Alu flanked by other repeats		–	–	22	–
Ya5NBC230	AC000100	GACAAAGAAAATGTCACAAGGGTAA	GGAAAGAATTATCTAGGACAGCTTTG	60	–	19	418-99
Ya5NBC231	BA422A16	Alu flanked by other repeats		–	–	22	–
Ya5NBC301	AC007682	TCATGCCTGAACATCTGCAT	ACCTACAGCTGTGCCTACCA	60	FP	2	795-677
Ya5NBC302	AL035665	CCTGCATACCCACCACATACC	GGCAGTCAGCTTTGACCTC	65	FP	20	395-72
Ya5NBC303	AL0136295	CTCCTCAAGGTCCCATGTTC	GGTGCCTCTGGGAATGAGTA	62	FP	14	426-111
Ya5NBC304	AL132642	GAGCTACTGGCACCTTCCAC	TTTTGACTCACCCTGCTTTTT	60	FP	14	368-60
Ya5NBC305	AP000966	CACATGGAGCTGTTTGCTGT	TGAGGGTTCTGTGAGAATTCAA	62	R/R	21	493-190
Ya5NBC307	AL133289	TCCCTGAAACAAAACCCATT	AAGACAAACCCCCATACA	65	FP	1	450-144
Ya5NBC308	AL133404	CAACAGAGAAGAAATGATCAGTGG	TGGGCCCTATATTTGAACAGA	60	FP	6	429-147
Ya5NBC309	AC020663	CCTCTACCTGCTGGGTTCAA	CCCAGGGACTCTCCAGAAA	55	FP	16	425-114
Ya5NBC310	AC008372	ATTGCAAATTGGCGATGTTC	CACCACTGAAGCATGCTAGG	62	IF	14	535-207
Ya5NBC311	AC008843	TCTTGGCAAGGAGATGTGAA	AATCACATCCGAGGGTGTCT	60	FP	5	584-279
Ya5NBC312	AC01069	CACTTGCCATTGACTTCCAA	GGCCTCTGGTTTCAATTGTC	60	IF	3	365-54
Ya5NBC313	AL121823	CACTTGCCATTGACTTCCAA	GGCTGGGTTGTGTGAGTTCT	60	FP	X	481-174
Ya5NBC314	AC016025	GTTCCAGGGGGGAAATGAAAT	GTGGGGCACTGTGTGATTC	60	FP	22	392-70

Table 1. (*continued*)

Name	Accession	5′ Primer sequence (5′–3′)	3′ Primer sequence (5′–3′)	A.T.[a]	Human diversity[b]	Chr.[c] Loc.	Product size[d] Filled-empty
Ya5NBC315	AP000474	GTAGACACCGCAGGCAACTC	AAAAGGATCCGTAAGAAGGAGA	62	FP	21	444-134
Ya5NBC317	AL132985	CCAAGTCAGGCACCAATAG	GATGGATAACCTTTTCCTGGT	60	FP	14	384-64
Ya5NBC319	AC007395	TTGCTGGTCCACAAACCATA	CCTTTGTCATCATGGTGCTG	60	FP	2	358-77
Ya5NBC320	AC009498	CCATCTCCCTCATTATTGTTCA	CCATTTGGGAGAAGGTTCAA	60	FP	2	478-161
Ya5NBC321	AL121748	GGAGATCCTTCTTTTTCAGCAA	GGAGGTGTCATCCTGGTACA	60	FP	10	455-145
Ya5NBC322	AL132800	AGTGCGTCAGATCCTGTTCA	GGGTCTTTGAAAAGTTCATGG	60	FP	14	451-129
Ya5NBC323	AC007076	TTGAAAGAGGAAGCCCAAGA	TCTCTGCTCCCCAACTCTTC	60	FP	7	556-268
Ya5NBC324	AC008268	TGTCTCAAGGGTCATCCTCA	TCCCCATCCCTAACTCTTTCTT	60	IF	2	486-164
Ya5NBC325	AC009479	CTTCTCTCTGAAATGCCAAT	CAGTTGAAAGGTTTGACAATACACC	60	IF	Y	501-184
Ya5NBC326	AL133500	CCAAGAGACCACTTCCTATTTCA	AATGGGGAGAGGACAGTCT	60	FP	X	539-216
Ya5NBC327	AL132799	AGGCAGGTTCAATGTTCAAA	TTGTCTTATTGTGCTGGCTAGA	60	IF	6	668-339
Ya5NBC329	AL121892	TTTTTCCCCTGTAGTTGGACA	TTGTTCAGGAGAGGGGAGGGA	60	FP	20	465-154
Ya5NBC330	AL133399	ATGCTGTGGGTTGCTAAGGA	CTGTCCCTTGTTTGGCTTGT	60	FP	11	402-88
Ya5NBC331	AL121593	TTCATGGCGAAAGGTTGATA	AGCTCCTGGCCAGATTAACA	62	FP	20	414-92
Ya5NBC332	AL050342	TGGAAACAGAGCAAGGTTGGACA	ACACAGGTCCTTGAAATATGAGC	65	FP	1	631-416
Ya5NBC333	AL117356	GGCATGCTATCATTCCCAAA	CCAAACTTCTGTTTGAGAGAATACG	60	IF	14	588-281
Ya5NBC334	AL132708	ACACTGTCTTGGAGGCATTC	CCTCCATCCCAGTACCATGA	60	FP	14	435-117
Ya5NBC336	AC007151	AGGCCCACATCACTGTAAGG	TGATCCATAGCTCTTTTTGTGC	60	FP	16	486-172
Ya5NBC338	AC009510	TCAAGAAGCTAAAGGCACCAA	AGGGGAAGAGGAAAAGATGAC	60	FP	12	564-271
Ya5NBC340	AL109985	TCCATATCCCTTGTCTGGTTC	CCTGACCAGGTCCAAATGAC	60	FP	14	468-145
Ya5NBC341	AC007899	ATGCAATTGCTGAAACACCAG	GGTGGACCCGAGATTTTCTTTC	60	FP	2	494-174
Ya5NBC342	AL049823	TTTTCCACAAATGGCACTGA	TGTCTGTGGCTCGTCATTTC	60	R/R	6	604-285
Ya5NBC343	AC005660	GACCACACTGGTCAGGGACT	CCCTCTTTGGTCTTGAGTGG	60	FP	10	457-154
Ya5NBC344	AL109853	CGTGAGAAAGCATAGGCAAC	TCCTTTCCTTATGCCTGCAA	60	FP	Xq	472-158
Ya5NBC346	AL098776	GGAGAACTAGTGTGGGAGCAG	ACACTCCCCTGTCCATTCCT	60	–	1	396-60
Ya5NBC347	AL035411	CATGCCATTGCTTTACGTT	TGGGGTAGATGGACTCATCC	60	IF	1	465-140
Ya5NBC349	AC011504	TCAAGAACGTGTGGCGCAAAT	GGATGTTGTCAGCAGCCAT	60	HF	19	469-53
Ya5NBC351	AP000459	TTCCTCCCCTTTTTCCTGTT	TGTCAGTATGTAAACCCATGCT	55	IF	21	437-123
Ya5NBC353	AL034549	CCATGTAACCTGGTAGACCTTT	GTTCAGCCGGGAACAGTGAGT	60	FP	20	432-119
Ya5NBC354	AC008039	GTAGCTTGGCCTGTGCTCTT	CCTCTGGGCTGAGAAACTCTT	65	IF	7	466-148
Ya5NBC355	AL078477	CATCTCACTTGAAAGCCCATT	TGTGTCTTAATGACCCTGGAAA	60	FP	11	802-481
Ya5NBC356	AF130343	CAGGGTCCTGTGAATCCAAT	GGAGACAGAAGAAAAGGGGAGA	62	FP	8	389-84
Ya5NBC359	AC007564	GCAAGTCCTATGCAAGGTGTT	AGGCTTTCAAGCCAGTGTT	60	FP	12	775-457
Ya5NBC360	AL031121	GAAACAAACATTTGGTAATGATGC	GACCAATGTCACTTATGAAATCCTT	60	FP	6	407-61
Ya5NBC361	AC007270	AATATTTTCTCCCATTCTTTTGG	TGTTAAAGCGCAAGTCACAA	60	IF	7	423-131
Ya5NBC362	AL050308	CAAGTTTGTTGGCATAGAGGTG	ATCAATCCAGGAGCCGTTTT	60	R/R	X	506-187

[a] Amplification of each locus required 2:30 minutes at 94 °C initial denaturing, and 32 cycles for one minute at 94 °C, one minute at annealing temperature (A.T.), and one minute elongation at 72 °C. A final extension time of ten minutes at 72 °C was also used.

[b] Allele frequency was classified as: fixed present (FP), low (LF), intermediate (IF), or high frequency (HF) insertion polymorphism. Fixed present: every individual tested had the Alu element in both chromosomes. Low frequency insertion polymorphism: the absence of the element from all individuals tested, except for one or two homozygous or heterozygous individuals. Intermediate frequency insertion polymorphism: the Alu element is variable as to its presence or absence in at least one population. High frequency insertion polymorphism: the element is present in all individuals in the populations tested, except for one or two heterozygous or absent individuals. (–) Indeterminable. (R/R) Repeat in repeat.

[c] Chromosomal location determined from Accession information or by PCR analysis of NIGMS monochromosomal hybrid cell line DNA samples.

[d] Empty product sizes calculated by removing the Alu element and one direct repeat from the filled sites that were identified.

other only 4 myr ago, then we can calculate that 6.64/10.64 (62 %) and 6.6/10.6 (62 %) of the Ya5 and Yb8 Alu elements should also be found at orthologous positions within the genomes of non-human primates. If we shift the divergence of humans and African apes to 6 million years ago then the estimates change to 4.64/10.64 (44 %) and 4.6/10.6 (43 %). However, less than 0.42 % of the elements were also located in orthologous positions in the genome of the common chimpanzee. The observed distribution of Ya5 and Yb8 Alu repeats located within the common chimpanzee genome would require a human and non-human primate divergence of greater than 10 myr ago. This is clearly a much older divergence time than is commonly accepted.

Three potential explanations may account for this. One is the selective removal of Alu elements from orthologous positions in non-human primate genomes effectively resulting in an ascertainment bias against elements in the non-human primate genomes because our elements were obtained by scanning a database of human genomic sequences. However, we consider this to be highly unlikely, because there are no known mechanisms to specifically remove Alu elements from primate genomes and even when an element is partially deleted from the genome it leaves behind a signature of itself.[33] A second and more likely explanation is that the amplification rate for these subfamilies has increased recently in the human lineage. Alternatively, the higher average ages for each of the Alu subfamilies than those previously reported may reflect a higher sequencing error rate in the genome database, resulting in an inflated age estimate for the Alu subfamilies. The estimated ages of the subfamilies are also inflated by the faster accumulation of non-CpG based mutations (as a result of the larger number of potential target sites) as compared to CpG nucleotides. Therefore, the use of the CpG-based mutation density for Alu subfamily age estimates will be much more accurate than the use of non-CpG mutation density-based estimates using the current draft sequence of the human genome. The magnitude of the putative sequencing errors can be estimated by comparing the previously reported non-CpG mutation density for these Alu subfamilies of approximately 0.4 % for the Ya5 and Yb8 Alu elements to the levels reported here of approximately 0.8 % for the same subfamilies. Therefore, the maximum possible error rate would be estimated as 0.8 % − 0.4 % = 0.4 %. In our data analysis, there are a few Alu elements with much higher mutation densities than previously seen. We are not sure whether these represent a small number of authentic, highly divergent subfamily members (approximately 10 % divergence), or the concentration of sequence errors in a few elements. Thus, other than the possibility of a few areas where errors may be concentrated, there is a relatively low sequencing error rate across the entire database, demonstrating the reliability of the draft human genomic sequence. Large scale re-sequencing of the Alu elements characterized in this paper would resolve this issue and allow for an accurate estimate of sequencing error rates within the draft human genomic sequence; it would also provide a refined estimation of the average age of the Alu Ya5 and Yb8 subfamilies as well.

SINE retroposition is the primary mode of mobilization of Alu elements, where mutations in the source gene(s) create their sequence evolution. However, previously we reported that gene conversion and genetic instability might have also significantly impacted the Alu sequence architecture.[18] Our analysis of the Yb8 mosaic elements also suggests that gene conversion may have influenced the evolution of the Yb8 Alu subfamily. Among the alternative explanations for the occurrence of mosaic elements, multiple parallel mutations seems unlikely; unless there was selection for these specific mutations, such as the post-transcriptional selection previously proposed.[8] However, a selection process that would only select for these specific mutations would be improbable. Recombination may have generated some of these mosaic elements, but multiple recombination events would be required, making it unlikely. Therefore, we believe gene conversion to be the most likely explanation for the existence of the mosaic Alu elements.

Our analysis of the human genomic diversity associated with the Ya5 and Yb8 Alu elements reported here resulted in the recovery of 106 new Alu insertion polymorphisms. The percentages of Alu insertion polymorphisms recovered from each subfamily were 25 % and 20 % for the Ya5 and Yb8 subfamilies, respectively. The percentages of Alu insertion polymorphisms in these two subfamilies are in good agreement with previously published insertion polymorphism estimates for these Alu subfamilies.[21] We can also estimate the total number of Alu insertion polymorphisms within the draft sequence of the human genome using our copy number estimates and the percentage of Alu insertion polymorphisms associated with each family. Using this approach we should recover 2640×0.25 or about 660 Ya5 Alu insertion polymorphisms and 1852×0.20 or about 370 Yb8 Alu insertion polymorphisms through the exhaustive analysis of the draft sequence of the human genome. Therefore, the exhaustive analysis of the entire Ya5 and Yb8 Alu subfamilies from the draft sequence of the human genome should generate a little more than 1000 Alu insertion polymorphisms from these subfamilies.

Additional Alu insertion polymorphisms that are present in diverse human genomes may also be recovered using PCR based display approaches such as those previously reported for Alu and LINE elements.[17,34] Each of the Alu insertion polymorphisms in the genome is a temporal genomic fossil that is identical by descent with a known ancestral state.[35,36] Previously, the analysis of Alu insertion polymorphisms has proved useful for the

Table 2. Alu Yb8 accession numbers, locations, human diversity, oligonucleotide primers and PCR parameters

Name	Accession	5' Primer sequence (5'-3')	3' Primer sequence (5'-3')	A.T.[a]	Human diversity[b]	Chr.[c] Loc.	Product size[d] Filled-empty
Yb8NBC1	AL049798	TACCAAGAGGATGTAAAACACAAGG	GGAACCCAAGGCTTATAATTTAGTC	60	FP	1	495-174
Yb8NBC2	U91327	GTAACTCTGTGGGTCCTCATATTATCACT	ATACCTCATCAGCAATAGGCAATAG	60	FP	12	461-130
Yb8NBC3	AC004804	AGATGCGCAAGTCCCTGATA	ATTTTTGGATTCAGCCAACG	61	HF	12	558-236
Yb8NBC4	AC005156	AGTGAGATGGTGGTTGCACA	AAAAACCTAAAGAGGGCAGT	60	FP	7	451-133
Yb8NBC5	AC004027	AAGGTCTAAGCGCAGTGGAA	TGTATGCAGGTTGCTTGCTC	60	LF	7	503-167
Yb8NBC6	AC006150	ATACCAAGACATCACACTGC	TAAAGCTGACACATTGTTGG	60	FP	7	606-203
Yb8NBC7	AC005048	AGGTTCATCCATGTTGTAGG	CTTAGAAGGGAATCCAGGAG	60	HF	7	605-285
Yb8NBC8	Z98950	AAGAAAACTGATGGGGAAAG	CCAACTAGAGAAACGGAGAA	60	IF	X	599-198
Yb8NBC9	AC004825	GTCCCCACCAATCCCTATCT	TGCTCAAAGTCCCACAGCTA	55	IF	14	655-322
Yb8NBC10	AC006352	CACGACAACACGTTTACCTCA	TTTCCTTTCAGGAACGTGGA	60	IF	7	505-165
Yb8NBC11	AL022477	TGGAAAGTGCGTGCCTTAAT	ACCTGAGGGAGAGACATTTCC	60	FP	6	510-188
Yb8NBC12	AC0006241	CCCAGCCAGGGTTTATTCTT	ACCCTGAAATGTCCTAGTGC	60	FP	9	487-160
Yb8NBC13	AC002331	TCTGGGTTTCTCTGGTGGAC	CTGGCAAATGCTACCCAAGT	60	LF	16	510-168
Yb8NBC14	AL08633	AGGAGACATTACAACTGATACTGC	TTGGCCTATTCCAGTCATGG	58	FP	X	499-167
Yb8NBC15	Z82211	CTAATTCCCCTGGCTGCATA	CTCTGGTACGGCCATAAAGC	60	FP	X	481-165
Yb8NBC16	U51244	TCACGAGAGGCCACTTTAGG	CCGACCACAAGCCAGAGTAT	60	FP	2	519-211
Yb8NBC17	AL049642	AGCATACAATTTGGCAAGCA	GCAGGAAGTGATTGTGCTGA	60	FP	X	507-184
Yb8NBC18	AL078476	TGTGGTGTGGGCTAGAGGATG	TGGGACTCAGATTTCTGATAGGA	56	IF	21	431-133
Yb8NBC19	AC007077	TTGGCATGTGAATTGTCTGAG	AAACCGTCAGTTGGGATCAG	55	FP	2	516-191
Yb8NBC20	AC002059	TCTGGGCAAGTCACTCAAAA	ACTGTTTCCATGGGCATGAT	56	R/R	22	502-158
Yb8NBC22	AC004875	CCACCAATGTCCTTTCTTAC	AGTGGAATCAGTCAATTGGT	50	FP	7	598-275
Yb8NBC23	AC005821	CCTCCTTGAGAAGGCAACAG	TGCATCTCCAGGTGTTCATC	60	FP	17	499-177
Yb8NBC24	AC005772	AACACTCCACCCAATAACTG	CATCTCTTCTTTACCTGTTTCC	56	HF	17	661-338
Yb8NBC26	AP000036	GGGTCCCTCTCTGAAGGTAAA	TTTTTCCTGCCAACCAACAC	56	HF	21	499-150
Yb8NBC27	AL021940	TTTTGGGAGGGACACAGTTC	GGGCCACACAAATAAATTTCC	55	FP	1	522-188
Yb8NBC28	AC003998	CTTTTTGATCATGAGCTGTG	ACAACAGAAACACCAGCTTT	60	FP	5	515-197
Yb8NBC29	AL09177	CTCTCCAGTATGGACAGAGG	AGTGCCCAGAACAGATATGA	60	FP	6	519-200
Yb8NBC30	Z95124	TTGCCTTGGATGGCATATCT	AAATGGCCGGAGTAAGTCCT	55	IF	X	497-194
Yb8NBC31	AC005046	TGAGGCAATTTGGAGGAGAG	TGGTGTGCTGTATGTTTTCCA	60	FP	7	518-188
Yb8NBC33	AP000352	GGAGAGCAAGCACACATAGT	GCACAATAGCAGAGGAGAAT	55	FP	22	496-167
Yb8NBC34	AP000168	ATGACCTTTGCATTTCACCA	GTTCAAGCCCCATCACATCT	60	FP	21	513-227
Yb8NBC35	AP000171	CCTTTTTGTCTCCTGGCTGA	CCTGAACACACCAGAGCAGA	60	R/R	21	497-220
Yb8NBC36	AP000156	CTAAGGAGAACGGGACAGAT	TATATTCTGGATCCCAGTCC	60	IF	21	506-183
Yb8NBC37	AP000193	AGAATCGAGTTTTCCGAGCA	GGGATGGTACAGATGGCATT	60	FP	21	486-162
Yb8NBC38	AC002367	CGAGAGAAAGGGGTAGAAAGC	AATGCCTTCCAAGGACATCTT	60	FP	X	480-311
Yb8NBC39	AP000111	CATGTGTCCATGGTGGTCAG	TCACCCCACTTCGGATTAAC	60	FP	21	456-129
Yb8NBC40	AP000080	TCCCAGGAGAAGGATGAGAA	CCTGTGTAGTTTTGGGCAATG	60	FP	8	500-171
Yb8NBC41	AC006011	TGCTTAAGGTTGGTCAGCAG	CAACCAGGAATGCTGTTTTACA	60	R/R	8	482-153
Yb8NBC42	AC005971	TTCCAGCATGTGAGAACATA	ATGCATCTTAGTCTGCTTGG	55	R/R	17	517-189
Yb8NBC43	AC004228	GTTGGACTGGGCTCTACCAG	GACAGCAGTTGCCATCTCA	60	FP	11	521-201
Yb8NBC44	AC005521	ACCACCCCTGGTACCAATTT	ATGCCTGCTCGTGCTTTACT	55	R/R	7	503-180
Yb8NBC45	AB020874	TGGCTTCAAGAGCATATCCA	GAGAAATGAACGCATTTTGG	55	FP	9	567-274
Yb8NBC46	AL031224	TGCATTTCTAAGCAGTACCAGTG	TCCATTCATTCGTGCCTTCT	60	FP	6	523-203
Yb8NBC48	AC005495	AGGGAACAGGTGATGTTTGG	GAGGATGCAAAAGCATGTGA	58	IF	17	511-178
Yb8NBC49	AL031904	GCAGTGGATTGGTTTTTCTG	GCTGAAAGAGGCATTGAAATC	60	IF	6	542-205

Name	Accession	Primer 1	Primer 2	Temp	Code	No.	Size
Yb8NBC51	AC005585	GCAGATAAGGGGTAACTGGA	GAAACACTTGGAAGATGCAA	60	FP	22	522-185
Yb8NBC52	AC004984	GGGGTCAGGGGAGGTAGAGAC	ACATCGGTGATTGGAAAACC	60	FP	7	515-196
Yb8NBC53	AL050338	GGTGGTTGCTAAAAGCTAGGG	GTGGCAGGTTTTGAGAGGAT	60	R/R	6	499-186
Yb8NBC54	AF165147	TGCCTACCTTTTGCACTTGA	AGCATATCATGACTGGGTTGAA	60	FP	21	454-220
Yb8NBC55	AC006374	CAAGGGGGCTATGCACTTTA	CCACCCAAATCTTTTTGTCA	60	FP	7	494-146
Yb8NBC56	AC007056	TGAACCACATGGACATCAAA	TATGCCCACCCAGATAATCC	50	FP	14	510-405
Yb8NBC57	AC006230	CCTTCACCTGCTTCCTTCAG	TATTCACAACAGGCCGCTTT	60	FP	14	519-191
Yb8NBC58	AC007319	ACACATCCCTGGGGCTTATT	CTTCCACTTGAAGTTTTCACTG	60	FP	4	566-251
Yb8NBC59	Z99758	GCAAAGGCCATTGAGAGAAA	AGGTCAAGCTGATCACTCACAG	60	FP	1	527-211
Yb8NBC60	AC004765	TGGGATGTGGTTCTGTGATT	TCAGACTGTGTTACTACGGGTCA	55	FP	12	528-304
Yb8NBC61	AC005040	GGAACTGCCTGAATGAACAA	CCTCTGCGCCCCTCCTACTAT	60	FP	2	539-209
Yb8NBC62	AL031368	TGCCACACATTGTTCTAGGC	TGCCAACTATTGGAGGAGATG	45	-	X	548-307
Yb8NBC63	AC005901	CTCCCCACGACAGAGAATTT	GATGGGTGGCAATCAGAGAT	60	FP	17	609-282
Yb8NBC65	AL031228	ATCTCATCTCCCTGCCTCTG	GGGAGGTCTGGAGATCTGTG	60	IF	6	517-186
Yb8NBC66	AC004981	TTTTAGGAATTGCCCCATTG	TCACTAATGTCCCAGCCAATC	55	FP	7	485-167
Yb8NBC67	AL022170	TCTCTACCCAGCTTTACCAA	GAGGACCAGCTTAGTTTGTG	55	IF	6	503-375
Yb8NBC68	AC004103	AGAAAAGTGCAAAGTGCCTA	TCCAAGCTCCTTAGTGTAGC	60	FP	X	518-188
Yb8NBC69	AC004458	CTTTCCCAAGGCCTCTGTC	GCCACACACCTAAAGCCATC	55	FP	7	557-242
Yb8NBC70	AC002456	CCAAAATCTTCCCAAGCAAA	TGTCAGTTCCTGGTTGGCTA	55	LF	2	541-208
Yb8NBC71	AC006222	GTGCTGAGCGCCAATCTTGA	GGACCAAGGGGTGCATTTA	55	-	7	493-167
Yb8NBC72	AC002432	GGTAGTGAAGAGGGCGAGGT	ATGCCAACTGGGTCTGCTAC	61	LF	7	526-210
Yb8NBC73	U47924	AGCCATTTTTCCCACTTCTGT	TGACCTCCCTTCAGGAATTG	45	R/R	12	437-111
Yb8NBC75	Z68328	CCCACTGTGTTTATTGTTCC	GCTAAAGTACCCAGACCAAG	55	FP	X	519-200
Yb8NBC76	AL031770	TGCAGTTAAAGCAACGCAAA	TTCACCTGACGAAGCCTGTA	60	FP	6	515-197
Yb8NBC77	AC007542	CGGAATGTTCTGAGGATCAAA	GGAAGCTCTGCACAACTCCTA	55	R/R	12	547-218
Yb8NBC78	AC005866	GAAAACGCCTCCACCATAGA	TCAAACCTGGCTACACACAGA	55	LF	12	536-215
Yb8NBC80	AC006249	ATTTCACAGTGCCCTGTCCT	TCCAGGCGATGAATTGACA	60	FP	18	456-123
Yb8NBC81	AC004837	TCCATTACTGAACACTTGGA	GAAGGGCAGTTTTGTGATAC	60	IF	7	502-173
Yb8NBC82	AC023806	TCATTGTTCTCTCAGTGTGTG	ATGGAAACAGTGTGAGGAAC	60	FP	6	566-239
Yb8NBC83	AC004741	CTTCCCCCTTCAGTGCCTAT	AACATGCATCCTGTGCAGAC	60	FP	7	528-199
Yb8NBC84	AL008628	TGGTCTGCAGGTTCCTCCTCT	TCAGCAAATAAAGCCCAAGG	60	FP	6	517-206
Yb8NBC85	Z98744	CACCGCGTGTTGATAGAGTC	CCACCTCACCATTCCAGAGT	60	FP	6	483-163
Yb8NBC86	AC004595	CTCCCTTTATCCCGGATGTT	GGAAGGCCATGGTGAAGATA	60	FP	7	460-353
Yb8NBC87	AC003048	TACTGAGCGCATCGAGGAAC	ACCGGCAGAGCGATAAATCAG	60	FP	X	537-210
Yb8NBC88	AC002524	CAGCTAGGCCTTGGAGATCA	TCAGAACCATGTTCTTGGAATC	60	FP	X	526-199
Yb8NBC90	AF025503	AGCTTTCTGCAAGCAAAGGA	AAAACAGTCTCAGGGCCAGT	55	R/R	3	488-153
Yb8NBC91	AC000083	AGCAACAGGGATTTGTCTGC	CAGTGGCTGACACACACACA	60	FP	22	513-194
Yb8NBC92	Z82189	CCTACCAAGACCGACGAGGAA	GCAGGGCTGCACTTTTTATC	60	FP	22	533-183
Yb8NBC93	AL035461	AAGTGAGTCCCAGGGCCTTCT	CACACAGGCACTTGTTTGGT	60	IF	20	601-274
Yb8NBC94	AL049633	GGGTGTGATGAAGGAAATAA	GCACTGCTCTGACCTCTATC	60	FP	20	601-274
Yb8NBC95	AL031346	GTGCTCAGCCAAAATGTCA	GGCACCAGGTTTCCTGAAGT	60	FP	22	532-210
Yb8NBC96	AC006222	GAACCCAAATAGCCAAAGCA	TGTCCTTCCCCCAGTTTATG	60	LF	22	498-224
Yb8NBC98	AC006441	TTGGCTATGTGAGTTAGATTGG	GGGCAATTTCAATAAGCAAGAG	60	FP	17	514-192
Yb8NBC99	Z92542	GATTCCAGTGCCTTCTGCTC	CAACCCATACAGTGCTTGGA	60	FP	X	524-200
Yb8NBC100	AB019438	CGCAGGTGCAATAATATGGA	AATTCCCTCCAGGTTGCTGT	60	IF	14	574-247
Yb8NBC101	AC005104	TTTGGCATCCACAGATTTGA	CTGTGCAGATTCCCCAAGTT	60	FP	?	593-273
Yb8NBC102	AL035088	GCCTGCAGAATGGAAGAAAC	CCAACTGAGTCCCCAGAAAA	60	HF	X	512-323
Yb8NBC103	AL022576	GCCTTTAAAGCCTAGCTCTC	GTGAATGCAAGTGAATGAAA	60	FP	X	531-205
Yb8NBC104	AC004638	GGGAGGAGTGGATGAATAAT	AAAACTGATGTCCAGCCATA	60	FP	16	533-207
Yb8NBC105	AF064865	TCACAGCACAATTCACAACTG	CTGGGTTGCATTTCATGGTA	60	IF	21	558-233

Table 2. (continued)

Name	Accession	5' Primer sequence (5'-3')	3' Primer sequence (5'-3')	A.T.ᵃ	Human diversityᵇ	Chrᶜ Loc.	Product sizeᵈ Filled-empty
Y8NBC107	AC006222	GTTTGGTTTTTCCGCAGTGT	GACTCGTCACTGGGTTGGAG	60	FP	4	527-207
Y8NBC108	AF164343	TGTCACTTGATTGTCCGCATA	TCAATGGCATCCTGAAAACA	60	IF	Y	550-194
Y8NBC109	AC006371	GTGCAACTTCAGTTTCTGCTAAGAT	CATGGTTATCTGCAAAGACTATGAC	55	IF	Y	532-212
Y8NBC110	AC006383	AATAGGCTGAATGCCCCAAT	CTAGCATTGCAATCCCTGCTTT	60	LF	Y	507-186
Y8NBC111	AC007320	CCAGTGTCATCATCCAGACTTATTC	TACACACACACATGCATTCTAAG	60	FP	Y	531-192
Y8NBC112	AC006999	GCATCTTAACCTAAATACCTGATGC	CAGGGACATAGGGTGTGAGTTACTA	60	FP	Y	503-192
Y8NBC114	AC004617	GGGTGAGATAGCTTAAGGAAAGAGA	AGATCTTCCCAAGAGAGCCTTTC	60	FP	Y	510-164
Y8NBC115	Dj102d24	TCATTCAGCCAACACTGACC	CAGGTTTTACCTCTCACCCTTGG	60	FP	22	628-297
Y8NBC117	Z82189	CAACCACACAGCCTAAAGACAT	GGCTGCACTTTTTATCCACCTA	60	FP	22	461-111
Y8NBC118	AC006548	GCAGAGACACATAAGACTGATTGAA	ACCTGGCTATGACCTGATAAATA	60	FP	22	519-200
Y8NBC119	Z95114	AGACCTTTGTCAGATGGATAGATTG	GTTTTGTGCTGTAAGGCTGAGTAG	60	FP	22	425-110
Y8NBC120	AC004019	CAGTGGATCTCCATTTTACCTCTC	GGAAAGGTTTCAGGAAGAAAGTG	60	IF	22	532-212
Y8NBC123	AL031846	TTTGGATGTTTGTTCCCTCT	GGTGAGGAGCAGAACACGAG	60	-	22	732-412
Y8NBC125	Dj309l22	AGCCAGAAACCCTGAACAAG	AAAGGCCCAGAAGTATACCA	60	IF	22	415-97
Y8NBC126	AC002055	AAAATGTCCCCTTTGTCCTTC	CCTACGCAGAAAACACCCTAGA	60	LF	22	438-118
Y8NBC129	AC004052	CCCAAACCTCCTAGATCTGC	CCCTGATTTCTTTCAGCAGTG	55	R/R	4	528-136
Y8NBC131	AC002994	TGTGGGTCTATTTCTGACTCCA	TCTACAAAACCGAAGCTGTT	55	FP	17	506-264
Y8NBC132	AC002458	GTTTCTGTGGGTTGGGATTC	AGCCAGCAAGAAACCTGAGTC	60	FP	4	507-187
Y8NBC133	Z84470	GCCATTGATCCCACAGAAAT	GCTGTGAATTCGTTGGTCCT	55	LF	4	536-232
Y8NBC134	AC002067	TGAGCAAAGGATTTGAATAGGC	AGGGTTCCAGTTTCCCCATA	60	LF	7	526-206
Y8NBC135	AC007392	TTCCTCCTCTTCTGGGACAA	GGAACCAAGGAGCAAAGAGA	60	FP	2	659-206
Y8NBC136	AC007055	CTTGCTCACACTCTGGTGGA	CTGATTTCACCGGTTTTTCTTC	60	-	14	530-196
Y8NBC137	AL031782	GGGTAAGTGGACAGGCGAAA	TGAAGCTATCTGGACCAGGAGT	55	FP	6	454-126
Y8NBC138	AL031653	GTTCCTTTCTTCTTCCTCAAAAG	TGCCTTTAATGTGCCATCTT	60	FP	20	650-332
Y8NBC140	AC006012	TTTACTGGACAGTTTGAAGC	CAGAAATGGTTCCTGTGTT	55	R/R	7	494-180
Y8NBC141	AC003950	GGGTAAGACAATAGTGGGGGATT	TTCACTAGATGTGCAAGGGGTTC	60	FP	17	530-219
Y8NBC142	AL049869	TCCAGTGCCTCAGAAGGTG	CATGGTGTCCTTCCTGTGTG	60	FP	14	487-162
Y8NBC143	AC009044	GGCTCTCTAAGCTAAGACAATCAA	CGTGCTCAAGGTATTGGTCA	55	FP	16	443-133
Y8NBC144	AL033531	TCACAGCGTGTGCATTACAA	AGGACTTCATTTTGGGGGATT	60	FP	1	578-255
Y8NBC145	AL035089	TGGTCCAGAACCTTCTCCAA	CAGGAACATGGGCTGAGTGT	65	FP	20	520-197
Y8NBC146	AC009028	CTCTTCTCTCCAGGAAACGTC	GGAGCTCTGCCTTACACTCAA	60	IF	16	887-592
Y8NBC147	AC010340	GAAATCTGTGCCATACGACGAAA	TGTGTGTGCACCATTTACA	55	FP	5	516-149
Y8NBC148	AC010582	CCAGGCCTCCATCTTTGATA	TCACTTTTGGGCATGTCAAG	60	IF	14	537-218
Y8NBC149	AL135746	TGAGTGAGTTCAGAAAAATCAAGG	TGATTAATTTACTTCATTTGGCAGT	60	FP	14	460-138
Y8NBC150	AP000855	CTGGCCATAAATTCCCTCAG	TCAGAAACTGCCCAAGAGAGA	55	-	21	474-160
Y8NBC151	AP000456	GGCACCAGGAGGAGAGAGAT	TGGTACCAAACTGCCTTCCT	60	FP	21	464-138
Y8NBC152	AC007911a	TGATGTTGACTTTGGCTTGA	GCTCCATAACTGGGTTCAGG	60	FP	18	520-183
Y8NBC153	AL049776	GGGAGTTAATCACTGTTCCTCAA	CAGGCTTTAGAATAAGCAGTGAGA	60	FP	14	569-248
Y8NBC154	AF172277	GGCCTGAGCACTGGTAGTTT	TGTGACTGGCCTATTTCACG	60	FP	7	469-147
Y8NBC155	AC010169	GGGAAGAGGGTCCAAGTGA	TTCCCCTCTACTCCCTCATTC	60	FP	3	421-90
Y8NBC156	AP000566	ACTCAGGCCCTTTCATTCTGC	ACTGGCAAGGAATGTGAGA	60	FP	21	554-234
Y8NBC157	AL121748	TATGGTTCTCAGCCATCACG	ATTCTTCCCCAAAGGGAGTC	60	IF	10	712-423
Y8NBC158	AC007671	GCAGAATACACCAAGCTGAGG	TGCCTGACTGTCCTATTTCAGA	60	FP	12	394-69
Y8NBC159	AC007680	TCACATTGTCCCTTCCTCAGC	TATGCAGGGCCTTCAACATA	60	FP	2	448-112
Y8NBC160	AC007284	CCACACATGGGTACCAGTCC	TTGCTTACCCACCAGTCACCTC	60	IF	Y	404-72
Y8NBC161	AC007100	CCATGTTCCAGGAAGTGTTCA	CACGCAAAGTTAACAGAAGTGC	60	FP	2	418-85
Y8NBC162	AL132987	GCACTCATTTTAGTGCTGCT	TGCAGTTCAGCCCCATAACAG	60	-	14	504-185
Y8NBC163	AL035467	TTTTCAATGCCTCTGTGCTG	TCCACTCACAAAGCTTCACG	60	FP	6	462-133
Y8NBC164	AC009509	TGACAACATCCGTGACAGAAA	TCACAGGCCCAATAAAACAT	60	FP	12	387-76
Y8NBC165	AC010200	TGGGATGAAGGGAAGATTGT	AACAGTGCCAATTCCTGAGAA	60	FP	12	465-151

Name	Accession	Forward Primer	Reverse Primer	Ta	Chr	Type	Size
Yb8NBC166	AL121852	CTGCTGCCTTCCCTAGACTG	CTCACTCTTAAGTGAACAGACTCAA	55	14	FP	570-248
Yb8NBC167	AL049777	CCTCTGGCTCCACAGGTAAA	ACTGGGTGCTTCAAAAGTGG	60	14	FP	415-95
Yb8NBC168	AL078603	AGATGCCCCCACATATCAAA	ATGGCATTCGTGGGTTTCTA	60	6	FP	469-150
Yb8NBC169	AC006480	TTTGGTAGCACTTCCGGTCT	GCCTCCTCACCCAATAGGA	60	7	–	426-99
Yb8NBC170	AL109653	TCCCCAAAGAAGGAGAGACA	TTCCCCATTCCACAATTTA	60	X	FP	599-275
Yb8NBC171	AL096771	TGCGTGTATTTTTCAACTGGTC	GACAAAGGGAAAATCCCATC	60	6	FP	537-206
Yb8NBC172	AC010197	GACACTTGAGTTTACCAGGAAAGA	CACTTCTAAATGAGACTGGCTTGAC	60	12	FP	408-88
Yb8NBC173	AC007250	GTAACTTCTGTCCCCCTTAAAATGT	CATGACCAGAATCACAGTCTCAA	60	18	FP	424-103
Yb8NBC175	AC011493	GCTAGGACCTCCAGTATTATGTTGA	AGCATTCCCTCTGTGAACTGAAAT	55	19	–	420-87
Yb8NBC177	AP000561	GACTACTCCAAAACTGCAAACAAAG	CTCAGTGAAATGCAAACTCTTTGAC	55	21	FP	474-150
Yb8NBC178	AL080286	TGGTTCTTCTAGGCTGCTTATTAC	TAGGTCCATTTCTCACCCTTTATAC	60	1	–	489-108
Yb8NBC181	AC007917	CATGTACCTTAGAATTCCACTCTCA	CCCCAAAGTTTATAGTCTGTGTTGTCT	55	3	FP	487-151
Yb8NBC183	AP000497	GGAAGAAATGCAAACTAAATAATGAGAG	CATTGTTACCAGCAACTTATTTACA	55	3	FP	465-140
Yb8NBC184	AP000495	AACTAACATAGCCCTGGTACAGAAA	CATTCCTGGATTACATCCTGTTTTA	55	3	FP	509-179
Yb8NBC185	AC008040	CACTTTGAAATAGTGCAAGGAATTT	CTCATTGACTCCTTTGACTCTTGTG	55	3	FP	500-211
Yb8NBC186	AC008055	ACAGTGGATGCTCCATATTTTACT	AGGTCTTGGAACTAGGAGCTTTATG	55	12	FP	503-177
Yb8NBC187	AL031905	GTCCATTCCATTTTACTGCTTACTC	TCCTGCATGTTTAACTTAACTTTCC	55	6	FP	491-179
Yb8NBC189	AC007684	GGAAGATTTGAGAGTGAAATACCC	ACATCATGGCCTGAACTAGTTTTTC	67	2	IF	541-220
Yb8NBC191	AL078604	AGTGACCAGAAAGCTCACAGTGTAT	CAGGGTTGCATGTACTGAGATATAG	65	6	FP	687-346
Yb8NBC192	AC006325	CTGCTCTACCCTAGGCTCTTCTATC	GCTCCTCTGCTTTTATGTGTTCTAC	55	6	HF	423-132
Yb8NBC193	AL049836	AGTGTTGTATTTAGGTCGGTGCAA	GCATGCTTGCAGGTGAGTC	55	7	–	536-164
Yb8NBC195	AL109733	CCTTTCTGGAAGGTTTTCAATG	CATGATGGAGAGGTACAAAGAGATT	58	14	R/R	531-201
Yb8NBC198	AL035695	AGGTCTCAAGTAGGATCCAGAGAAG	GTTTGTCGAACGTGGGAAGTTA	58	X	FP	528-194
Yb8NBC199	AC007377	TAGATGGCTTTAGCAATTATAAGGT	CAATTTCAGGAACACTGTAAAGTCA	60	6	FP	848-522
Yb8NBC200	AC008041	GGAAAGCAGATCTTCTGACTCCTA	AGGCCAATTACGGAATACATAACTC	55	2	FP	426-99
Yb8NBC201	AC007558	GGAGAAAATGTAAGGTTTCTAGCAC	ACCAATGCAACTATCTCACTGACA	60	3	IF	476-145
Yb8NBC202	AC006984	ATGTAGAGAAAGCTGGTCTGTGAAG	CATTTCCTATCTTACTCTCCATGTC	58	7	FP	405-90
Yb8NBC203	AL031655	CAAGATTGTCAGTGACCCTTAAGAA	TCATTCTAACCCGTTCAGATGTACT	55	7	FP	518-200
Yb8NBC204	AC004885	CTTCCTCTTTTCCTATTCAAGCTCT	CAGAAGAAAGTGCATCTCTCAAAAG	55	20	FP	489-181
Yb8NBC205	AC007543	CTGCCCTAAAACTCAGTGACTAAAA	AAGTGGACCTGAAACCTATGTGATA	60	7	FP	419-93
Yb8NBC206	AL033525	CTATTTCCTAGCGTGCCTGAGAT	TGAGGTGATTTACCTTCACTCTACC	60	12	FP	486-153
Yb8NBC208	AP000243	ATCACTAAAGAGACTGTTGGCGTTT	TAATCTAGGCAAAACTGCTTACCC	60	1	IF	357-111
Yb8NBC209	AC006511	AAGTCATTGCTTACAGAAAACTGGAG	CCATGGAATGACATCTAGGTTGTT	60	21	FP	548-227
Yb8NBC210	AC007198	TGACGTGCAGACTACCTAATGTAAA	TACTTTTAGAACAGGGCCTCAGAAT	60	12	FP	416-91
Yb8NBC211	AC007165	TGAAAACCAGTTGCCAGAA	GGGGCTAACTCAGATGTCCA	55	?	–	383-58
Yb8NBC212	AC006561	TGGACTACAACATACCATCCTCA	CAGCGTGTTGTGACATTTGTT	55	18	FP	418-102
Yb8NBC213	AL033381	CAGCATTTGTGCCTTATCCTT	TTGGTGTTCTTGAAGAGGTGAA	60	12	FP	562-210
Yb8NBC214	AC000159	CCCTGCAAACCATTTCATCT	GGGTGAGAGGGCTGTTAGAA	60	6	FP	719-472
Yb8NBC216	AC005999	TCTTTGTTCCTATCTTACCCAATTC	AGGCACAAAGTGGAAACTGG	55	11	FP	400-84
Yb8NBC217	AC005988	CTTGCCATAGCCCTTTTTGT	GGGTCTTCTTGGGGATGAA	50	7	FP	648-308
Yb8NBC218	AC005099	TGAGGTGAGGCTCTGTTCC	CTGTTTCTTTTCCTGCACCA	60	17	–	531-215
Yb8NBC219	AC004866	TACCACATTGCCTCCACATC	GCACATGGCAACTGTCTGAG	62	7	FP	580-231
Yb8NBC220	AL024509	AAAGAGGTTTCTTTGGCTGGA	AAACTCACTGAATGGCTGACAC	60	7	FP	387-65
Yb8NBC221	AL034370	AATTCAAGCCAATGAACCAC	TCAGTGCTCTGAAGAAGCTCA	60	6	FP	431-97
Yb8NBC222	AC005552	AGCTCCCACTCCCTACTTT	GGGAGAGTTCAGATGGGAAA	60	X	FP	426-102
Yb8NBC223	AC004915	CCATCCACAAATATCACAAAGC	TGGGGAACCATACCTTCTTG	60	17	FP	550-226
Yb8NBC224	AC004861	GGTCACTGTATTTTCCTCAAATCC	GGTGTTTGAGTATATGTAGGTGTGC	58	7	IF	417-102
Yb8NBC225	AC005868	GAGTCCAGCCCATTTTAGCA	CCCAGCACAAACATGTCATT	60	12	FP	449-135
Yb8NBC226	AC004853	GGAAATGCAAATGCCAATA	CATGATGGTTTGCTGCAACT	60	7	FP	537-189
Yb8NBC227	AC005799	AAGAAAAGGGAAGCCTGGAG	CAGTCATCACCAGCCATGAG	60	17	IF	881-546

Table 2. (*continued*)

Name	Accession	5′ Primer sequence (5′-3′)	3′ Primer sequence (5′-3′)	A.T.[a]	Human diversity[b]	Chr.[c] Loc.	Product size[d] Filled-empty
Yb8NBC228	AC005722	GTGTCCAGACCTGTGGCTCT	CCAGACAGCTGGGGTTTT	62	FP	17	630-310
Yb8NBC229	AC005754	CCCAGTTTCTACTTTGCACTG	TGCCAACTGAGCACTTCTTG	60	FP	5	411-90
Yb8NBC230	AB014460	CAAATGGCCGTGTTCTTTT	GTGTCCACGGATCTTTGCAG	62	IF	16	458-124
Yb8NBC231	AC005618	GGAAGACTCCCTTGTTCAGG	ATGCATTATTTTCCCCACA	60	FP	5	501-181
Yb8NBC232	AL023875	TGTGAATCCCACAGTCAGAAA	TTCACAGCTGGATCAGTTCAA	55	FP	X	402-83
Yb8NBC233	AC004702	TCCACATGGATGGAAGATGA	GTGGTCTGCAAGGGAACAGT	60	FP	17	446-114
Yb8NBC234	AC005207	ACCTGCAAAAGAGGCGTAGA	CTAATGAGGCCACCACTCAA	60	FP	17	523-198
Yb8NBC235	AC005221	CATTCTGGGCACCTCACTTT	CCATCCAAATTGCCTAAGGT	60	FP	5	473-146
Yb8NBC236	AL021939	CTGCTTTCAGTGTCCAGGAT	CAAAGCCTATGTCTCGCTCA	60	FP	6	775-449
Yb8NBC237	AC004613	GCCAAAATCAACTGCCAAAC	TGCTGAGGATAGAGCTATAGCCAGA	60	IF	7	491-164
Yb8NBC238	AC004592	AATGAAGTCACCTGCCCTTG	CCTGAAGAGATGGTGGAAGG	60	FP	5	437-117
Yb8NBC239	AF031078	TTGCTGACAGATCAGGGATG	TCCCCCTTCAAACCTATTCC	55	FP	X	730-419
Yb8NBC240	AC004452	TTCACAGTGATTCGTGCTCTCA	GGTGTCTTCTGAGAAATGCCTA	60	FP	7	564-257
Yb8NBC241	AC004391	GGACTGTGTCTAAGGGTGTCCT	GGTAATTGGGAGCAGTTGAGA	60	IF	7	424-93
Yb8NBC242	AC002349	ATCCACCATCAGGGAATCAA	TGCAGATCTTATCAGCACATTG	60	FP	X	450-117
Yb8NBC244	AF043945	CGGATGTCCCTTTACCACAT	CACTGCGTGGTTCATCACTT	60	FP	21	403-77
Yb8NBC245	AC004029	AACCCATTGTCTCATGTCTAGC	CTCTCATCCAACAAAAGTCAGTGT	60	FP	7	647-318
Yb8NBC246	AC002981	CACCACCTTTCAACCAGGAA	ATCGCTGGAATGTGGTTCTC	60	FP	X	464-149
Yb8NBC247	AC002366	GCACGCAAAGTAGTGCGTTGG	TGCACCCACTTGATATGCTT	60	FP	X	551-259
Yb8NBC248	AC003088	TTTCTTTCCCTCTCGCATGT	CCCTTTGGTCTCGACACATT	60	FP	7	441-120
Yb8NBC249	298049	ATGGGCCCAAATAAAAGGAT	GTGATGGCCTTGACAGCAT	62	FP	6	491-148
Yb8NBC250	AC002462	GGGATCCCAGACATTGATTT	TTGCTCCTCACTTGCTCCTT	60	FP	?	397-63
Yb8NBC251	AC002477	CGGCCCTGATATGTCTTTGA	TCCACAAGGCAAATGGATA	60	FP	X	838-500
Yb8NBC252	AC002123	GCCCACCATCGAGATCTACT	TCCACATCCATCAGAGCTT	62	FP	5	424-107
Yb8NBC253	AF001548	CAAAGCACTTAATGCCAATCA	CCCCTCCTTCTCCTTTGCTA	60	-	16	473-144
Yb8NBC254	AC002088	GGGGGAAACATTACTACAGAGG	ATATATTTTTGGCCAGGTACGG	55	FP	7	740-413
Yb8NBC255	AC000062	GGAATGAAGTGTCCACAGATGA	CAGAGGCAGGGAGACCAG	55	FP	13	423-93
Yb8NBC256	Z73986	CCCACAATTTCCACTTCAGG	GCATTGCTTCCCTTCTATTTC	55	FP	X	503-24
Yb8NBC257	269921	CTGCACCAAAAGAGACACACA	GCAAAAACATAGAAAGCGGTGT	55	FP	4	508-187
Yb8NBC258	AC009429	TTAGTGGTTCCTGCATGTGG	AGCGCAGGGTTAGTAGCAAA	60	-	?	431-103
Yb8NBC259	AC015600	TCCAGGAAAAGGGAACATT	TTTCAAGAGAAGGGCAACA	60	-	?	547-227
Yb8NBC260	AC012000	TTTCCACCATCAGTCCCTCT	AGGGACTTAGGAGTGATTTTAGTG	55	-	2	661-327
Yb8NBC261	AC009478	GCAGCACTTAATGCCAATCA	TCATCGTTCTTTAGCTCCTCTG	60	-	2	375-50
Yb8NBC262	AC020728	ATCCAGATTTGCAGGACCAC	CCTCAGCTAAGTGCCAGGAG	60	FP	5	687-361
Yb8NBC263	AC009318	GAAAGAGAGGGCAGCATTGT	AAAGTTTATGCTCCCGCTGA	60	-	12	518-177
Yb8NBC264	AC007619	AAGCAAGACATATGCATGAAAAA	ATCGTTTTAAATGTTGCATACCA	60	FP	12	781-475
Yb8NBC265	AC025436	TGCCAACTGACACTTCTTG	CCCTTTTGGATTCTCTCTGC	60	FP	5	434-114
Yb8NBC266	AC009078	TATTCATTGCCTCCCTTGA	ATGCTCCAACCCTTTAGG	60	FP	16	475-149
Yb8NBC267	AC008925	GGGAAGTTTCAACAAAACCAGA	CACCACTGAATGATACCCTTTT	60	FP	5	777-447
Yb8NBC268	AC016681	TGGGATAGAGGAAGAAGAGACAA	CCTTTTCATCCAACTACCACTG	60	IF	Y	517-188
Yb8NBC269	AF241735	CACGCTTAACCTCTACCACCA	TGGACTCCCACTGAGATGTG	60	-	X	587-261
Yb8NBC270	AC007489	CTTCTGCAGCTCCTGACTGA	AGTCTAGGCTTCGGATGCAG	55	-	16	403-72
Yb8NBC271	AC023602	GGAAAAACTGCATGCTAGGC	CAGTGAATGTTTCCCTGTGGT	60	FP	3	493-168
Yb8NBC272	AC023602a	TGCAGAATGTTTGTTCTTGGAG	TTTCCCTAGCTCCTTGAAATG	60	-	3	537-215
Yb8NBC273	AC011284	CTCTTTGTGGGGGAGAAG	CATGCTCCTTGGGAACTCTC	55	R/R	7	431-112
Yb8NBC274	AL133305	TCAACATCAAACCCACTGAA	TCCGAGGGAGGAATGAGATA	55	FP	14	489-168
Yb8NBC275	AL122000	TTCCTGAAAAAGACCTACACCTG	TTTGGCCTTATGTGACAAGC	60	FP	1	472-146
Yb8NBC276	AL109928	TCTGCTGGGGTCAGAAAAAC	GGCGGTTGTTAAGTGGAAA	55	-	20	507-183
Yb8NBC277	AL121944	AAACAATGAACTGAAGGGGACT	CCCTGCAGCCCTGTATAAATCA	60	-	6	404-82

[a] Amplification of each locus required 2:30 minutes at 94 °C initial denaturing, and 32 cycles for one minute at 94 °C, one minute at annealing temperature (A.T.), and one minute elongation at 72 °C. A final extension time of ten minutes at 72 °C was also used.

[b] Allele frequency was classified as: fixed present (FP), low (LF), intermediate (IF), or high frequency (HF) insertion polymorphism. Fixed present: every individual tested had the Alu element in both chromosomes. Low frequency insertion polymorphism: the absence of the element from all individuals tested, except for one or two homozygous or heterozygous individuals. Intermediate frequency insertion polymorphism: the Alu element is variable as to its presence or absence in at least one population. High frequency insertion polymorphism: the element is present in all individuals in the populations tested, except for one or two heterozygous or absent individuals. (−) Indeterminable. (R/R) Repeat in repeat.

[c] Chromosomal location determined from Accession information or by PCR analysis of NIGMS monochromosomal hybrid cell line DNA samples.

[d] Empty product sizes calculated by removing the Alu element and one direct repeat from the filled sites that were identified.

Table 3. Alu Ya5 subfamily associated human genomic diversity

Elements	African American					Greenland natives/Asian[c]					European					Egyptian					Avg. Het[b]
	Genotypes					Genotypes					Genotypes					Genotypes					
	+/+	+/−	−/−	fAlu	Het[a]	+/+	+/−	−/−	fAlu	Het[a]	+/+	+/−	−/−	fAlu	Het[a]	+/+	+/−	−/−	fAlu	Het[a]	
A. *Intermediate frequency*																					
Ya5NBC5	2	5	5	0.38	0.49	3	2	8	0.31	0.44	1	6	11	0.22	0.36	2	8	4	0.43	0.51	0.45
Ya5NBC22	3	15	1	0.55	0.51	4	14	0	0.61	0.49	1	16	1	0.50	0.51	19	1	0	0.98	0.05	0.39
Ya5NBC27	0	5	14	0.13	0.24	0	8	11	0.21	0.34	2	7	9	0.31	0.44	2	7	10	0.29	0.42	0.36
Ya5NBC35	9	10	1	0.70	0.43	5	12	2	0.58	0.50	8	12	0	0.70	0.43	7	13	0	0.68	0.45	0.45
Ya5NBC37	2	2	13	0.18	0.30	1	4	12	0.18	0.30	3	2	15	0.20	0.33	4	3	10	0.32	0.45	0.34
Ya5NBC45	7	7	2	0.66	0.47	19	0	0	1.00	0.00	17	0	0	1.00	0.00	8	3	0	0.86	0.25	0.18
Ya5NBC51	4	10	3	0.53	0.51	5	6	8	0.42	0.50	6	7	7	0.48	0.51	3	8	9	0.35	0.47	0.50
Ya5NBC57	10	1	2	0.81	0.32	4	8	3	0.53	0.52	13	2	1	0.88	0.23	9	1	1	0.86	0.25	0.33
Ya5NBC61	10	6	3	0.68	0.44	5	2	10	0.35	0.47	9	7	1	0.74	0.40	8	4	5	0.59	0.50	0.45
Ya5NBC96	17	2	0	0.95	0.10	9	5	3	0.68	0.45	18	1	0	0.97	0.05	16	3	0	0.92	0.15	0.19
Ya5NBC102	3	2	13	0.22	0.36	0	0	6	0.00	0.00	3	4	12	0.26	0.40	2	0	13	0.13	0.24	0.25
Ya5NBC109	7	11	1	0.66	0.46	7	11	2	0.63	0.48	5	13	1	0.61	0.49	7	8	4	0.58	0.50	0.48
Ya5NBC120	7	11	0	0.69	0.44	15	4	0	0.90	0.19	8	12	0	0.70	0.43	14	5	0	0.87	0.24	0.32
Ya5NBC123	5	7	7	0.45	0.51	6	5	4	0.57	0.51	14	5	1	0.83	0.30	11	5	1	0.79	0.34	0.41
Ya5NBC131	0	5	6	0.23	0.37	0	9	8	0.27	0.40	0	11	6	0.32	0.45	0	15	2	0.44	0.51	0.43
Ya5NBC132	4	0	5	0.44	0.52	9	0	0	1.00	0.00	13	0	0	1.00	0.00	11	0	1	0.92	0.159	0.17
Ya5NBC148	7	6	6	0.53	0.51	2	6	12	0.25	0.39	0	0	20	0.00	0.00	0	0	17	0.00	0.00	0.22
Ya5NBC150	17	0	0	1.00	0.00	4	0	14	0.22	0.36	19	0	1	0.95	0.10	17	0	1	0.94	0.11	0.14
Ya5NBC154	0	12	5	0.35	0.47	0	7	9	0.22	0.35	0	12	8	0.30	0.43	3	4	13	0.25	0.39	0.41
Ya5NBC160	2	7	9	0.31	0.44	0	0	19	0.00	0.00	0	0	19	0.00	0.00	0	4	12	0.13	0.23	0.17
Ya5NBC174	0	5	3	0.31	0.46	0	3	8	0.14	0.25	0	12	8	0.30	0.43	2	5	9	0.28	0.42	0.39
Ya5NBC182	2	9	9	0.33	0.45	9	8	0	0.77	0.37	5	6	7	0.44	0.51	1	10	3	0.43	0.51	0.46
Ya5NBC201	6	6	5	0.53	0.51	4	7	6	0.44	0.51	16	3	0	0.92	0.15	8	7	2	0.68	0.45	0.41
Ya5NBC210	0	4	15	0.11	0.19	0	1	15	0.03	0.06	0	4	16	0.10	0.19	0	4	12	0.13	0.23	0.17
Ya5NBC216	5	7	5	0.50	0.52	6	8	5	0.53	0.51	7	12	0	0.68	0.44	0	0	10	0.00	0.00	0.37
Ya5NBC219	0	10	9	0.26	0.40	1	12	7	0.35	0.47	0	11	9	0.28	0.41	0	0	6	0.00	0.00	0.32
Ya5NBC221	5	7	4	0.53	0.51	9	5	3	0.68	0.45	16	0	1	0.94	0.11	13	2	0	0.93	0.13	0.30
Ya5NBC311[c]	12	1	6	0.66	0.46	11	4	2	0.77	0.37	15	1	4	0.78	0.36	11	2	4	0.71	0.43	0.41
Ya5NBC313[c]	9	3	5	0.62	0.49	4	6	6	0.44	0.51	2	8	3	0.46	0.52	5	6	3	0.57	0.50	0.50
Ya5NBC324[c]	0	8	1	0.44	0.52	0	15	1	0.47	0.51	0	14	4	0.39	0.49	0	15	1	0.47	0.51	0.51
Ya5NBC325[c]	0	10	10	0.25	0.39	0	9	9	0.25	0.39	0	11	9	0.28	0.41	0	6	6	0.25	0.39	0.39
Ya5NBC327[c]	2	9	9	0.33	0.45	13	6	1	0.80	0.33	19	0	0	1.00	0.00	7	6	1	0.71	0.42	0.30
Ya5NBC333[c]	5	5	9	0.40	0.49	4	7	8	0.49	0.49	3	8	8	0.37	0.48	5	3	5	0.50	0.52	0.50
Ya5NBC347[c]	17	2	1	0.90	0.19	4	7	8	0.40	0.49	7	8	2	0.65	0.47	11	1	5	0.68	0.45	0.40
Ya5NBC351[c]	3	12	3	0.55	0.51	7	9	3	0.61	0.49	13	3	3	0.76	0.37	11	1	5	0.68	0.45	0.46
Ya5NBC354[c]	0	2	16	0.06	0.11	2	6	10	0.28	0.41	10	4	5	0.63	0.48	2	4	9	0.27	0.41	0.35
Ya5NBC361[c]	0	9	10	0.24	0.37	2	11	5	0.42	0.50	0	5	12	0.15	0.26	3	3	7	0.35	0.47	0.40

Locus	n1	n2	n3	Freq	Het	n1	n2	n3	Freq	Het	n1	n2	n3	Freq	Het	n1	n2	n3	Freq	Het	Avg. Het
B. *High frequency*																					
Ya5NBC16	20	0	0	1.00	0.00	20	0	0	1.00	0.00	10	10	0	0.75	0.39	20	0	0	1.00	0.00	0.10
Ya5NBC18	17	1	0	0.97	0.06	18	1	0	0.97	0.05	20	0	0	1.00	0.00	20	0	0	1.00	0.00	0.03
Ya5NBC98	17	0	1	0.94	0.11	18	2	0	0.95	0.10	17	1	2	0.88	0.22	16	3	0	0.92	0.15	0.14
Ya5NBC157	20	0	0	1.00	0.00	20	0	0	1.00	0.00	20	0	0	1.00	0.00	15	0	1	0.94	0.12	0.03
Ya5NBC212	16	1	0	0.97	0.06	20	0	0	1.00	0.00	20	0	0	1.00	0.00	18	0	0	1.00	0.00	0.01
Ya5NBC349[c]	19	1	0	0.98	0.05	14	2	4	0.75	0.39	18	0	0	1.00	0.00	12	0	2	0.86	0.25	0.17
C. *Low frequency*																					
Ya5NBC24	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0.00
Ya5NBC28	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0.00
Ya5NBC38	0	0	16	0.00	0.00	0	2	15	0.06	0.14	0	0	20	0.00	0.00	0	0	18	0.00	0.00	0.00
Ya5NBC54	0	0	14	0.00	0.00	0	0	12	0.00	0.00	0	1	8	0.06	0.11	0	1	7	0.06	0.13	0.09
Ya5NBC135	0	1	18	0.03	0.05	0	1	19	0.03	0.05	0	0	20	0.00	0.00	0	0	17	0.00	0.06	0.01
Ya5NBC147	0	1	17	0.03	0.05	0	0	20	0.00	0.00	0	0	20	0.03	0.00	0	1	17	0.03	0.06	0.04
Ya5NBC155	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	19	0.00	0.00	0	0	20	0.00	0.00	0.00
Ya5NBC171	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0.00
Ya5NBC172	0	0	20	0.00	0.00	0	0	18	0.00	0.00	0	0	20	0.00	0.00	0	0	18	0.00	0.00	0.00
Ya5NBC184	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	19	0.00	0.00	0.00
Ya5NBC194	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0.00
Ya5NBC197	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	19	0.00	0.00	0.00
Ya5NBC203	0	0	20	0.00	0.00	0	0	17	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0.00
Ya5NBC204	0	0	19	0.00	0.00	0	0	20	0.00	0.00	0	0	17	0.00	0.00	0	0	15	0.00	0.00	0.00
Ya5NBC214	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	5	15	0.13	0.22	0	0	19	0.00	0.00	0.06
Ya5NBC223	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	19	0.00	0.00	0.00

[a] This is the unbiased heterozygosity.
[b] Average heterozygosity is the average of the population heterozygosity.
[c] The following were tested using DNA samples from Asian individuals.

Table 4. Alu Yb8 subfamily associated human genomic diversity

| | African American | | | | | Greenland natives/Asian[c] | | | | | European | | | | | Egyptian | | | | | Avg. |
| Elements | Genotypes | | | | | Genotypes | | | | | Genotypes | | | | | Genotypes | | | | | |
	+/+	+/−	−/−	fAlu	Het[a]	+/+	+/−	−/−	fAlu	Het[a]	+/+	+/−	−/−	fAlu	Het[a]	+/+	+/−	−/−	fAlu	Het[a]	Het[b]
A. *Intermediate frequency*																					
Yb8NBC3	10	2	0	0.92	0.16	12	4	0	0.88	0.23	4	6	0	0.70	0.44	9	0	0	1.00	0.00	0.207
Yb8NBC7	5	8	0	0.69	0.51	4	14	0	0.61	0.49	1	16	1	0.50	0.51	19	1	0	0.98	0.05	0.39
Yb8NBC8	8	1	0	0.94	0.11	10	4	0	0.86	0.25	11	1	2	0.82	0.30	12	2	3	0.77	0.37	0.26
Yb8NBC9	3	5	10	0.31	0.44	2	3	13	0.19	0.32	5	1	9	0.37	0.48	0	7	8	0.23	0.37	0.402
Yb8NBC10	9	9	0	0.75	0.39	9	11	0	0.73	0.41	12	7	0	0.82	0.31	11	5	0	0.84	0.27	0.344
Yb8NBC18	1	0	15	0.06	0.12	2	9	9	0.33	0.45	0	6	14	0.15	0.26	1	6	11	0.22	0.05	0.22
Yb8NBC30	8	6	0	0.79	0.35	7	11	0	0.69	0.44	5	8	0	0.69	0.44	14	2	0	0.94	0.12	0.338
Yb8NBC36	5	14	1	0.60	0.49	8	0	0	1.00	0.00	10	9	0	0.76	0.37	8	8	0	0.75	0.39	0.312
Yb8NBC48	0	4	6	0.20	0.34	0	1	2	0.17	0.33	0	3	4	0.21	0.36	0	2	3	0.20	0.36	0.347
Yb8NBC49	1	9	10	0.28	0.41	7	8	5	0.55	0.51	5	9	6	0.48	0.51	1	8	9	0.28	0.41	0.46
Yb8NBC65	7	6	5	0.56	0.51	3	10	7	0.40	0.49	7	4	9	0.45	0.51	2	5	9	0.28	0.42	0.481
Yb8NBC67	8	5	5	0.58	0.50	9	6	4	0.63	0.48	13	2	0	0.93	0.13	4	7	4	0.50	0.52	0.406
Yb8NBC71	0	3	13	0.09	0.18	3	3	10	0.28	0.42	0	5	12	0.15	0.26	2	2	9	0.23	0.37	0.304
Yb8NBC77	2	2	16	0.15	0.26	2	0	16	0.11	0.20	0	1	17	0.03	0.06	0	0	16	0.00	0.00	0.13
Yb8NBC80	1	4	15	0.15	0.26	2	5	12	0.24	0.37	3	1	15	0.18	0.31	2	5	8	0.30	0.43	0.344
Yb8NBC93	1	3	10	0.18	0.30	7	5	2	0.18	0.30	7	2	5	0.57	0.51	12	4	1	0.82	0.30	0.35
Yb8NBC96	0	7	9	0.22	0.35	0	14	3	0.41	0.50	0	3	15	0.08	0.16	0	5	7	0.21	0.34	0.338
Yb8NBC106	4	6	7	0.41	0.50	2	8	10	0.30	0.43	0	2	18	0.05	0.10	3	5	11	0.29	0.42	0.362
Yb8NBC108	2	11	7	0.38	0.48	2	10	7	0.37	0.48	0	3	11	0.11	0.20	3	4	10	0.29	0.43	0.396
Yb8NBC109	0	11	8	0.29	0.42	1	11	8	0.33	0.45	4	1	6	0.41	0.51	7	0	11	0.39	0.49	0.467
Yb8NBC120	5	8	5	0.50	0.51	5	6	8	0.42	0.50	8	7	3	0.64	0.48	4	2	6	0.42	0.51	0.499
Yb8NBC125	0	0	20	0.00	0.00	0	3	16	0.08	0.15	0	3	17	0.08	0.14	0	5	14	0.13	0.24	0.132
Yb8NBC146	18	0	2	0.90	0.19	12	1	1	0.89	0.20	16	0	2	0.89	0.20	10	1	6	0.62	0.49	0.268
Yb8NBC148	11	0	2	0.85	0.27	11	1	6	0.64	0.48	6	2	10	0.39	0.49	13	3	4	0.75	0.41	0.411
Yb8NBC157	19	0	1	0.95	0.10	6	5	1	0.71	0.43	3	9	6	0.42	0.50	16	2	2	0.85	0.26	0.322
Yb8NBC160	0	12	8	0.25	0.39	0	13	7	0.33	0.45	0	10	10	0.25	0.39	1	6	13	0.20	0.33	0.387
Yb8NBC189	10	10	0	0.75	0.39	18	2	0	0.95	0.10	9	7	2	0.69	0.44	18	2	0	0.95	0.10	0.254
Yb8NBC201	5	9	0	0.59	0.50	3	8	7	0.39	0.49	9	5	6	0.58	0.50	2	9	8	0.34	0.46	0.488
Yb8NBC208	5	6	5	0.50	0.52	18	2	1	0.91	0.18	10	8	2	0.70	0.43	15	4	1	0.85	0.26	0.346
Yb8NBC225[c]	10	9	1	0.73	0.41	12	2	4	0.72	0.41	11	6	3	0.70	0.43	8	2	5	0.60	0.50	0.4375
Yb8NBC227[c]	10	8	2	0.70	0.43	5	6	5	0.50	0.52	18	2	0	0.95	0.10	15	4	1	0.85	0.26	0.326
Yb8NBC230[c]	1	2	11	0.14	0.25	0	0	19	0.00	0.00	0	2	15	0.06	0.11	1	4	3	0.38	0.50	0.217
Yb8NBC237[c]	13	4	1	0.83	0.29	12	5	2	0.76	0.37	15	2	0	0.94	0.11	10	8	1	0.74	0.40	0.293
Yb8NBC241[c]	0	0	16	0.00	0.00	2	0	14	0.13	0.23	2	3	10	0.23	0.37	1	6	8	0.27	0.41	0.25
Yb8NBC268[c]	0	13	5	0.36	0.48	0	7	12	0.18	0.31	1	9	8	0.31	0.44	0	5	12	0.15	0.26	0.37

																					[b]
B. *High frequency*																					
Yb8NBC24	12	2	0	0.93	0.14	13	1	0	0.96	0.07	15	0	0	1.00	0.00	12	0	0	1.00	0.00	0.052
Yb8NBC26	12	1	0	0.96	0.08	14	1	0	0.97	0.07	9	1	0	0.95	0.10	13	0	0	1.00	0.00	0.061
Yb8NBC102	16	0	2	0.89	0.20	20	0	0	1.00	0.00	20	0	0	1.00	0.00	18	0	0	1.00	0.00	0.051
Yb8NBC181	14	0	0	1.00	0.00	13	0	0	1.00	0.00	8	9	3	0.63	0.48	13	0	0	1.00	0.00	0.12
Yb8NBC192	20	0	0	1.00	0.00	20	0	0	1.00	0.00	19	1	0	0.98	0.05	20	0	0	1.00	0.00	0.012
C. *Low frequency*																					
Yb8NBC5	3	10	4	0.47	0.01	0	2	10	0.08	0.16	0	0	18	0.00	0.00	0	0	20	0.00	0.00	0.042
Yb8NBC13	15	3	0	0.92	0.08	0	0	17	0.00	0.00	0	0	15	0.00	0.00	0	0	12	0.00	0.00	0.019
Yb8NBC69	0	0	18	0.00	0.00	0	0	17	0.00	0.00	0	0	16	0.00	0.00	0	0	16	0.00	0.00	0
Yb8NBC100	0	1	17	0.03	0.06	0	5	15	0.13	0.22	0	1	18	0.03	0.05	0	0	0	0.00	0.00	0.083
Yb8NBC110	0	0	18	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0
Yb8NBC126	1	6	7	0.29	0.07	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0.018
Yb8NBC133	0	2	18	0.05	0.06	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0	0	20	0.00	0.00	0.016
Yb8NBC134	0	4	18	0.09	0.17	0	0	17	0.00	0.00	0	0	10	0.00	0.00	0	0	19	0.00	0.00	0.042

[a] This is the unbiased heterozygosity.
[b] Average heterozygosity is the average of the population heterozygosity.
[c] The following were tested using DNA samples from Asian individuals.

study of human population genetics.[35–43] The newly identified Alu insertion polymorphisms from the Ya5 and Yb8 Alu subfamilies should prove useful for the study of human population genetics.

Materials and Methods

Cell lines and DNA samples

The cell lines used to isolate primate DNA samples were as follows: human (*Homo sapiens*), HeLa (ATCC CCL2); and chimpanzee (*Pan troglodytes*), Wes (ATCC CRL1609). Cell lines were maintained as directed by the source and DNA isolations were performed using Wizard genomic DNA purification (Promega). Human DNA samples from the European, African American, Asian, Egyptian, and Greenland Native population groups were isolated from peripheral blood lymphocytes[44] available from previous studies.[18]

Computational analyses

Initial screening of the GenBank non-redundant and high throughput genomic sequence (HTGS) databases was performed using the Basic Local Alignment Search Tool (BLAST)[45] available from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). Copy number estimates were determined using Megablast and the draft human genome sequence database.[46] The database was searched for exact complements to the oligonucleotide 5′-CCATCCC-GGCTAAAAC-3′ and 5′-TGCGCCACTGCAGTCCG-CAGTCCG-3′ that are exact matches to a portion of the Alu Ya5 and Yb8 subfamily consensus sequences (respectively) that contain unique diagnostic mutations.[21] Sequences that were exact complements to the oligonucleotides were then subjected to more detailed annotation. A region composed of 500-1000 bases of flanking DNA sequence directly adjacent to the sequences identified from the databases that matched the initial GenBank BLAST query were subjected to annotation using the RepeatMasker2 program from the University of Washington Genome Center server (http://ftp.genome.washington.edu/c/s.dll/RepeatMasker) or Censor from the Genetic Information Research Institute (http://www.girinst.org/Censor_Server-Data_Entry_Forms.html).[47] These programs annotate the repeat sequence content of individual sequences from humans and rodents. A complete list of the Alu elements identified from the GenBank search is available from MAB. The copy numbers for each subfamily of Alu elements were determined by screening the draft sequence of the entire human genome with the oligonucleotides shown above.[23] For the Yb8 subfamily analysis, the database was searched for matches to the consensus Yb8 sequence without the seven-nucleotide duplication (287 bases). The sequences were then subjected to more detailed analysis using MegAlign (DNAStar version 3.1.7 for Windows 3.2) selecting only for Yb8 intermediate elements containing between one and seven of the Yb8 diagnostic sites.

Primer design and PCR amplification

PCR primers were designed from flanking unique DNA sequences adjacent to individual Ya5 and Yb8 Alu elements using the Primer3 software (Whitehead Institute for Biomedical Research, Cambridge, MA, USA) (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). The resultant PCR primers were screened against the GenBank non-redundant database for the presence of repetitive elements using the BLAST program, and primers that resided within known repetitive elements were discarded and new primers were designed. PCR amplification was carried out in 25 μl reactions using 50-100 ng of target DNA, 40 pM of each oligonucleotide primer, 200 μM dNTPs in 50 mM KCl, 1.5 mM MgCl$_2$, 10 mM Tris-HCl (pH 8.4) and *Taq*® DNA polymerase (1.25 units) as recommended by the supplier (Life Technologies). Each sample was subjected to the following amplification cycle: an initial denaturation of 150 seconds at 94 °C, one minute of denaturation at 94 °C, one minute at the annealing temperature, one minute of extension at 72 °C, repeated for 32 cycles, followed by a final extension at 72 °C for ten minutes. For analysis, 20 μl of each sample was fractionated on a 2% agarose gel with 0.25 μg/ml ethidium bromide. PCR products were directly visualized using UV fluorescence. The sequences of the oligonucleotide primers, annealing temperatures, PCR product sizes and chromosomal locations for all Ya5 and Yb8 elements can be found on our website (http://129.81.225.52). Phylogenetic analysis of all the ascertained Alu elements was determined by PCR amplification of human and non-human primate DNA samples. The human genomic diversity associated with each Alu element was determined by the amplification of 20 individuals from each of four populations (African-American, Greenland Native or Asian, European and Egyptian) (160 total chromosomes). The chromosomal location of Alu repeats identified from clones that had not been previously mapped was determined by PCR amplification of National Institute of General Medical Sciences (NIGMS) human/rodent somatic cell hybrid mapping panel 2 (Coriell Institute for Medical Research, Camden, NJ).

References

1. Smit, A. F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657-663.
2. Deininger, P. L. & Batzer, M. A. (1993). Evolution of retroposons. In *Evolutionary Biology* (Heckht, M. K., *et al.*, eds), pp. 157-196, Plenum Publishing, New York.
3. Maklalowski, W., Mitchell, G. & Labuda, D. (1994). Alu sequences in the coding regions of mRNA: a source of protein variablility. *Trends Genet.* **10**, 188-193.
4. Makalowski, W., Zhang, J. & Boguski, M. (1996). Comparative analysis of 1,196 orthologous mouse

and human full-length mRNA and protein sequences. *Genome Res.* **6**, 846-857.

5. Weiner, A., Deininger, P. L. & Efstratiadis, A. (1986). The reverse flow of genetic information: pseudogenes and transposable elements derived from non-viral cellular RNA. *Annu. Rev. Biochem.* **55**, 631-661.

6. Boeke, J. (1997). LINEs and Alus - the polyA connection. *Nature Genet.* **16**, 37-43.

7. Jurka, J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl Acad. Sci. USA,* **94**, 1872-1877.

8. 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.

9. Batzer, M. A., Kilroy, G. E., Richard, P. E., Shaikh, T. H., Desselle, T. D. *et al.* (1990). Structure and variability of recently inserted Alu family members. *Nucl. Acids Res.* **18**, 6793-6798.

10. Deininger, P. L., Batzer, M. A., Hutchison, C. & Edgell, M. (1992). Master genes in mammalian repetitive DNA amplification. *Trends Genet.* **8**, 307-312.

11. Deininger, P. L. & Daniels, G. (1986). The recent evolution of mammalian repetitive DNA elements. *Trends Genet.* **2**, 76-80.

12. Leeflang, E. P., Liu, W-M., Hashimoto, C., Choudary, P. V. & Schmid, C. W. (1992). Phylogenetic evidence for multiple Alu source genes. *J. Mol. Evol.* **35**, 7-16.

13. Leeflang, E. P., Liu, W-M., Chesnokov, I. N. & Schmid, C. W. (1993). Phylogenetic isolation of a human Alu flounder gene: drift to new subfamily identity. *J. Mol. Evol.* **37**, 559-565.

14. Schmid, C. & Maraia, R. (1992). Transcriptional and transpositional selection of active SINE sequences. *Curr. Opin. Genet. Dev.* **2**, 874-882.

15. Shen, M., Batzer, M. A. & Deininger, P. L. (1991). Evolution of the Master Alu Gene(s). *J. Mol. Evol.* **33**, 311-320.

16. Batzer, M. A., Deininger, P. L., Hellmann-Blumberg, U., Jurka, J., Labuda, D. *et al.* (1996). Standardized nomenclature for Alu repeats. *J. Mol. Evol.* **42**, 3-6.

17. Roy, A. M., Carroll, M. L., Kass, D. H., Nguyen, S. V., Salem, A-H. *et al.* (1999). Recently integrated human Alu repeats: finding needles in the haystack. *Genetica,* **107**, 1-13.

18. Roy, A. M., Carroll, M. L., Nguyen, S. V., Salem, A-H., Oldridge, M. *et al.* (2000). Potential gene conversion and source gene(s) for recently integrated Alu elements. *Genome Res.* **10**, 1485-1495.

19. Slagel, V., Flemington, E., Traina-Dorge, V., Bradshaw, H., Jr. & Deininger, P. L. (1987). Clustering and subfamily relationships of the Alu family in the human genome. *Mol. Biol. Evol.* **4**, 19-29.

20. Willard, C., Nguyen, H. T. & Schmid, C. W. (1987). Existence of at least three distinct Alu subfamilies. *J. Mol. Evol.* **26**, 180-186.

21. Batzer, M. A., Rubin, C. M., Hellmann-Blumberg, U., Alegria-Hartman, M., Leeflang, E. P. *et al.* (1995). Dispersion and insertion polymorphism in two small subfamilies of recently amplified Alu repeats. *J. Mol. Biol.* **247**, 418-427.

22. Deininger, P. L. & Batzer, M. A. (1999). Alu Repeats and human disease. *Mol. Genet. Metab.* **67**, 183-193.

23. Lander, E. S., Linton, L. M., Birren, B. & Nusbaum, R. *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature,* **409**, 860-921.

24. Arcot, S. S., Shaikh, T. H., Kim, J., Bennett, L., Alegria-Hartman, M. *et al.* (1995). Sequence diversity and chromosomal distribution of ''young'' Alu repeats. *Gene,* **163**, 273-278.

25. Labuda, D. & Striker, G. (1989). Sequence conservation in Alu evolution. *Nucl. Acids Res.* **17**, 2477-2491.

26. Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucl. Acids Res.* **8**, 1499-1504.

27. Miyamoto, M. M., Slightom, J. L. & Goodman, M. (1987). Pylogenetic relations of humans and African apes from DNA sequences in the pseudo-η-globin region. *Science,* **238**, 369-373.

28. Arcot, S. S., Wang, Z., Weber, J. L., Deininger, P. L. & Batzer, M. A. (1995). Alu repeats: a source for the genesis of primate microsatellites. *Genomics,* **29**, 136-144.

29. Maeda, N., Wu, C.-I., Bliska, J. & Reneke, J. (1988). Molecular evolution of intergenic DNA in higher primates: pattern of DNA changes, molecular clock, and evolution of repetitive sequences. *Mol. Bio. Evol.* **5**, 1-20.

30. Kass, D. H., Batzer, M. A. & Deininger, P. L. (1995). Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. *Mol. Cell. Biol.* **15**, 19-25.

31. Batzer, M. A., Gudi, V., Mena, J. C., Foltz, D. W., Herrera, R. J. & Deininger, P. L. (1991). Amplification dynamics of human-specific (HS) Alu family members. *Nucl. Acids Res.* **19**, 3619-3623.

32. Hutchinson, G. B., Andrew, S. E., McDonald, H., Goldberg, Y. P., Grahm, R. *et al.* (1993). An *Alu* element retroposition in two families with Huntington disease defines a new active *Alu* subfamily. *Nucl. Acids Res.* **21**, 3379-3383.

33. Edwards, M. C. & Gibbs, R. A. (1992). A human dimorphism resulting from loss of an Alu. *Genomics,* **3**, 590-597.

34. Sheen, F.-M., Sherry, S. T., Risch, G. M., Robichaux, M., Nasidze, I. *et al.* (2000). Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. *Genome Res.* **10**, 1496-1508.

35. Batzer, M. A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D. H. *et al.* (1994). African origin of human-specific polymorphic Alu insertions. *Proc. Natl. Acad. Sci. USA,* **91**, 12288-12292.

36. Stoneking, M., Fontius, J. J., Clifford, S. L., Soodyall, H., Arcot, S. S. *et al.* (1997). Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res.* **7**, 1061-1071.

37. Comas, D., Calafell, F., Benchemsi, N., Helal, A. & Lefranc, G. *et al.* (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.

38. 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.

39. Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M. E., Ricker, C. E. *et al.* (2000). The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66**, 979-988.

40. Majumder, P. P., Roy, B., Banerjee, S., Chakraborty, M., Dey, B. *et al.* (1999). Human-specific insertion/deletion polymorphisms in Indian populations and

their possible evolutionary implications. *Eur. J. Hum. Genet.* **7**, 435-446.

41. Perna, N. T., Batzer, M. A., Deininger, P. L. & Stoneking, M. (1992). Alu insertion polymorphism: a new type of marker for human population studies. *Hum. Biol.* **64**, 641-648.

42. Tishkoff, S. A., Ruano, G., Kidd, J. R. & Kidd, K. K. (1996). Distribution and frequency of a polymorphic Alu insertion at the plasminogen activator locus in humans. *Hum. Genet.* **97**, 759-764.

43. Watkins, W. S., Ricker, C. E., Bamshad, M. J., Carroll, M. L., Nguyen, S. V. *et al.* (2001). Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *Am. J. Hum. Genet.* **68**, 738-752.

44. Ausabel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G. *et al.* (1987). *In Current Protocols in Molecular Biology*, New York.

45. Altschul, S., Madden, T., Schaffer, A., Zhang, Z., Miller, W. & Lipman, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389-3402.

46. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203-214.

47. Jurka, J. P., Klonowski Dagman, V. & Pelton, P. (1996). CENSOR - a program for identification and elimination of repetitive elements from DNA sequences. *Comput. Chem.* **20**, 119-121.

Edited by J. Karn