Article No. mb981984 *J. Mol. Biol.* (1998) **281**, 843–856

JMB

High-resolution Cartography of Recently Integrated Human Chromosome 19-Specific Alu Fossils

Santosh S. Arcot[1], Aaron W. Adamson[1], Gregory W. Risch[2] Jake LaFleur[2], Myles B. Robichaux[2], Jane E. Lamerdin[1] Anthony V. Carrano[1] and Mark A. Batzer[1,2]*

[1]*Human Genome Center L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore CA 94551, USA*

[2]*Departments of Pathology Biometry and Genetics Biochemistry and Molecular Biology, Stanley S. Scott Cancer Center, Neuroscience Center of Excellence, Louisiana State University Medical Center, 1901 Perdido St. New Orleans, LA 70112, USA*

**Corresponding author*

The recently inserted subfamilies of Alu retroposons (Ya5/8 and Yb8) are composed of approximately 2000 elements. We have screened a human chromosome 19-specific cosmid library for the presence of Ya5/8 and Yb8 Alu family members. This analysis resulted in the identification of 12 Ya5/8 Alu family members and 15 Yb8 Alu family members from human chromosome 19. The total number of Ya5/8 and Yb8 Alu family members located on human chromosome 19 does not differ from that expected based upon random integration of Alu repeats within the human genome. The distribution of both subfamilies of Alu elements along human chromosome 19 also appears to be random. DNA sequence analysis of the individual Alu elements revealed a low level of random mutations within both subfamilies of Alu elements consistent with their recent evolutionary origin. Oligonucleotide primers complementary to the flanking unique sequences adjacent to each Alu element were used in polymerase chain reaction assays to determine the phylogenetic distribution and human genomic variation associated with each Alu family member. All of the chromosome 19-specific Ya5/8 and Yb8 Alu family members were restricted to the human genome and absent from orthologous positions within the genomes of several non-human primates. Three of the Yb8 Alu family members were polymorphic for insertion presence/absence within the genomes of a diverse array of human populations. The polymorphic Alu elements will be useful tools for the study of human population genetics.

Keywords: retroposon; insertion polymorphism; physical map

Introduction

The Alu family of short interspersed elements (SINEs) is a large family of mobile genetic elements that have expanded to a copy number of 500,000 within the last 65 million years of primate evolution (for reviews see Deininger, 1989; Okada, 1991; Schmid & Maraia, 1992; Deininger & Batzer, 1993, 1995). The expansion of the Alu family to such a high copy number has occurred as a result of the amplification of only a few active ''master'' or ''source'' genes producing a hierarchical subfamily structure for these elements. The elements may be divided into subfamilies or clades of related Alu elements that share common diagnostic mutations. Here, we will use the standardized nomenclature to refer to the various Alu subfamilies (Batzer *et al*., 1996b). Two of the most recently formed subfamilies of Alu elements have been termed Ya5 and Ya8. Ya8 Alu family members contain all five of the Ya5 diagnostic substitutions as well as three additional mutations such that we refer to this lineage of Alu sequences collectively as the Ya5/8 subfamily. Independently, and in an overlapping time frame, a second derivative of the Y subfamily designated as Yb8 that is character-

Correspondence address: M. A. Batzer, Department of Pathology, Louisiana State University Medical Center, 1901 Perdido St., New Orleans, LA 70112, USA.

Abbreviations used: BLAST, basic local alignment search tool; FISH, fluorescence *in situ* hybridization; LINE, long interspersed element; LLNL, Lawrence Livermore National Laboratory; PCR, polymerase chain reaction; SINE, short interspersed element.

E-mail address of the corresponding author: mbatze@lsumc.edu

0022–2836/98/350843–14 $30.00/0

ized by eight different mutations has also amplified within the human genome.

The amplification of Ya5/8 and Yb8 Alu family members to a copy number of approximately 2000 members is an ongoing process that has started in very recent evolutionary history (Batzer *et al*., 1995). Most of the members of the Ya5/8 and Yb8 subfamilies have amplified so recently in primate evolution that they are absent from the genomes of non-human primates although a few Ya5/8 Alu repeats are found within chimpanzee and gorilla genomes (see Deininger & Batzer, 1993, 1995 for reviews). Some of the Ya5/8 and Yb8 Alu family members have integrated so recently within the human genome that they are polymorphic for presence/absence within different human genomes (Batzer *et al*., 1994, 1996a; Stoneking *et al*., 1997). There are also examples of *de novo* Alu insertion events that give rise to neurofibromatosis and hemophilia (Wallace *et al*., 1991; Vidaud *et al*., 1993).

The insertion of Alu repeats within the human genome may occur as the result of random process, or may be biased based upon local sequence context or gene density. Human chromosome 19 is known to have a high G+C content and is also extremely gene-dense (Ashworth *et al*., 1995). Early fluorescence *in situ* hybridization (FISH) based studies using metaphase chromosomes suggested that the gross distribution of interspersed repeated DNA sequences throughout the genome was not completely random with some local areas of concentration of different classes of repeats (Korenberg & Rykowski, 1988). In addition, Korenberg & Rykowski (1988) also demonstrated that long interspersed elements (LINEs) preferentially integrate in A+T-rich regions (G bands) of the genome whereas Alu SINEs integrate in G+C-rich regions (R bands). Sainz *et al.* (1992) have also reported a non-random distribution of Alu and L1 elements along human chromosome 21. Recently, a common integration site motif for Alu repeats has been identified by Jurka (1997) suggesting the existence of an endonucleolytic activity for the integration of SINEs and LINEs. An endonuclease was simultaneously identified as part of one of the open reading frames from LINE elements by Boeke and co-workers (Feng *et al*., 1996). The existence of an endonucleolytic activity for the integration of SINEs and LINEs and a common integration site sequence both have potential implications for the dispersal of Alu interspersed repeats throughout the human genome. The recently inserted ''young'' Alu elements are a rich source of genomic fossil landmarks for the study of the dispersal of Alu repeats, phylogenetics and human population genetics. Here, we report the identification of recently inserted Alu repeats along an entire human autosome and discuss the implications of these observations with respect to the dispersal of these elements throughout the genome.

Results

Library screening and physical location

A total of 116 Ya5/8 and 75 Yb8 probe-positive clones were identified from approximately 10,000 arrayed chromosome 19-specific cosmid clones. Due to the redundancy of the cosmid library that was screened, the total number of probe positive clones is higher than the actual number of young Alu repeats. Since several of the probe-positive clones represent overlapping cosmids containing the same Alu element, the clones were organized into clusters using a graphical analytical tool (Browser) that interfaces with the cosmid map of chromosome 19. This process reduced the number of potential cosmid contigs containing unique Alu elements by a factor of four (189 to 49). Of the 49 cosmid contigs, 22 contigs contained Ya5/8 Alu family members and 27 contigs contained Yb8 Alu repeats. It is possible that the first round of screening may have identified several false positive clones due to the high density of the clones on the filter. For the second round of screening, a single positive cosmid clone from each one of the 49 contigs was subcloned into a plasmid vector and rescreened for the presence of the specific Ya5/8 or Yb8 Alu element. The subcloning resulted in the identification of 12 independent Ya5/8 and 14 independent Yb8 Alu family members; a total of 26 chromosome 19-specific Alu repeats. The positive clones from the first round of screening that were discarded after the second round were also tested by Southern blot analysis for the presence of Ya5/8 or Yb8 Alu elements.

Integration of the probe-positive cosmid clones into the chromosome 19 physical map (Ashworth *et al*., 1995) resulted in the determination of the physical location of the individual Alu repeats along human chromosome 19. The individual Alu elements from both subfamilies are dispersed along both arms of the chromosome as shown in Figure 1. Physical locations, cosmid clone numbers, subfamily designations and GenBank accession numbers for each Alu element are shown in Table 1. It is also interesting to note that the Alu elements from both subfamilies reside within G and R bands along human chromosome 19. Two previously reported Alu repeats from the Ya5/8 and Yb8 subfamilies (PV30B and LDLR) were not isolated during the library screening (see Discussion).

DNA sequence analysis

DNA sequencing of each Alu repeat was performed using internal Alu subfamily-specific primers described by Batzer & Deininger (1991) and Batzer *et al*. (1995). In the case Yb8 clones Sb19.2 and Sb19.7, the primer used to obtain the sequence of the 5′ flanking region did not perform well, possibly due to random mutations in the sequencing primer annealing sites. Since these Alu elements were positively identified using a

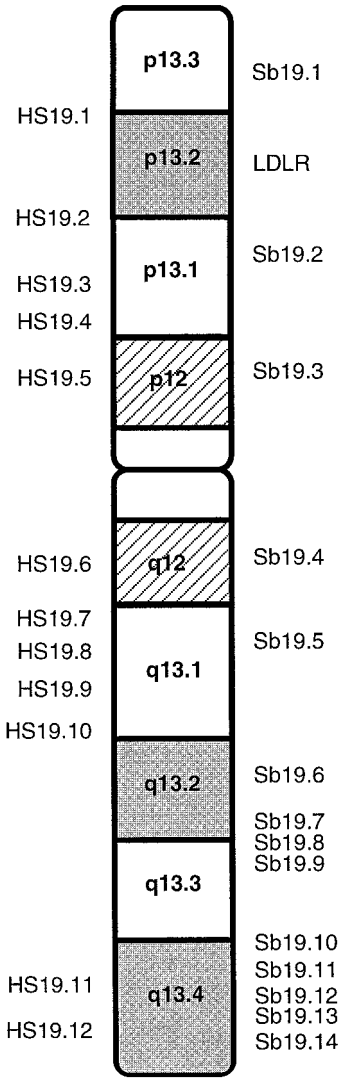

Figure 1. Idiogram of human chromosome 19 showing the approximate location of the Alu elements. The physical location of each Alu element was determined using the metric physical map of the chromosome as a framework to locate the cosmid contigs containing Ya5/8 and Yb8 Alu elements using subfamily-specific oligonucleotide as probes for hybridization to arrayed high-density cosmid clones. The physical location of individual cosmid contigs was previously determined by metaphase FISH analysis using specific cosmid clones from the contigs as probes for hybridization. The exact physical location of individual Alu elements can be obtained using the cosmid clone ID as a marker on the LLNL chromosome 19 website *via* the internet (see Table 1).

Table 1. Characteristics of chromosome 19-specific young Alu repeats

Name	Subfamily[a]	Cosmid clone ID[b]	Location[c]	Genbank ID
HS19.1	Ya5	F16706	p13.2/13.3	AF015147
HS19.2	Ya5	F15251	p13.1/13.2	AF015148
HS19.3	Ya5	F18633	p13.1	AF015149
HS19.4	Ya5	F19106	p13.1	AF015150
HS19.5	Ya5	F24757	p12	AF015151
HS19.6	Ya5	F16949	q12	AF015152
HS19.7	Ya5	F23468	q13.1	AF015153
HS19.8	Ya5	F22132	q13.1	AF015154
HS19.9	Ya8	F15702	q13.1	X54177
HS19.10	Ya5	F21419	q13.1/13.2	AF015155
HS19.11	Ya5	F19081	q13.4	AF015156
HS19.12	Ya5	F16107	q13.4	AF015157
PV30B	Ya8	–	–	U02532
Sb19.1	Yb8	F15546	p13.3	AF015158
Sb19.2	Yb8	F25413	p13.1	AF015159
Sb19.3	Yb8	F23467	p12	AF015160
Sb19.4	Yb8	F18993	q12	AF015161
Sb19.5	Yb8	F21682	q13.1	AF015162
Sb19.6	Yb8	F20428	q13.2	AF015163
SB19.7	Yb8	F24330	q13.2/13.3	AF015164
Sb19.8	Yb8	F23145	q13.2/13.3	AF015165
Sb19.9	Yb8	F17058	q13.3	AF015166
Sb19.10	Yb8	F18874	q13.3/13.4	AF015167
Sb19.11	Yb8	F20766	q13.4	AF015168
Sb19.12	Yb8	F22777	q13.4	AF015169
Sb19.13	Yb8	F24232	q13.4	AF015170
Sb19.14	Yb8	F22371	q13.4	AF015171
LDLR	Yb8	–	p13.2	L35531

[a] Subfamily nomenclature described by Batzer *et al.* (1996b).
[b] Additional information about the cosmid clones can be obtained through the internet (http://www-bio.llnl.gov/genome/html/chrom_map.html).
[c] Locations are based upon the physical (Ashworth *et al.*, 1995) and FISH maps (Brandriff *et al.*, 1994; Gordon *et al.*, 1995) of chromosome 19.

subfamily-specific hybridization oligonucleotide, a new sequencing primer was designed that is complementary to the probe region. The new sequencing primer is approximately 240 bases from the 5′ end of the Alu element and its sequence is 5′-CTCTGTCGCCCAGGCCG-GACTGCGG-3′. Flanking unique nucleotide sequences were used to design oligonucleotide primers for the Alu insertion polymerase chain reaction (PCR) assay. PCR products from each locus were TA cloned for finished sequencing of each Alu repeat.

The nucleotide sequences from the chromosome 19 Ya5/8 Alu family members are shown in Figure 2. Each of the Ya5/8 Alu family members were flanked by perfect direct repeats that ranged in size from nine (HS19.10) to 17 nucleotides (HS19.4, HS19.8 and HS19.11) with an average length of 13.2 nucleotides. The length and base composition of the direct repeats is similar to that previously reported for recently integrated Alu elements (Arcot *et al.*, 1995a, 1996; Batzer *et al.*, 1990, 1995) and is considered a hallmark of recent retroposition events. Each of the Ya5/8 Alu repeats also contained an oligo(dA)-rich tail that varied from 12 (HS19.4) to 30 (HS19.2) nucleotides in length. Excluding insertions and deletions a total of 30 out of 3363 nucleotides were different from the Ya5/8 subfamily consensus sequence. If we divide the mutations into those that occurred at CpG dinucleotides (12) and those at non-CpG positions (18), the level of non-CpG substitutions can be used as an indication of the age of these Ya5/8 Alu family

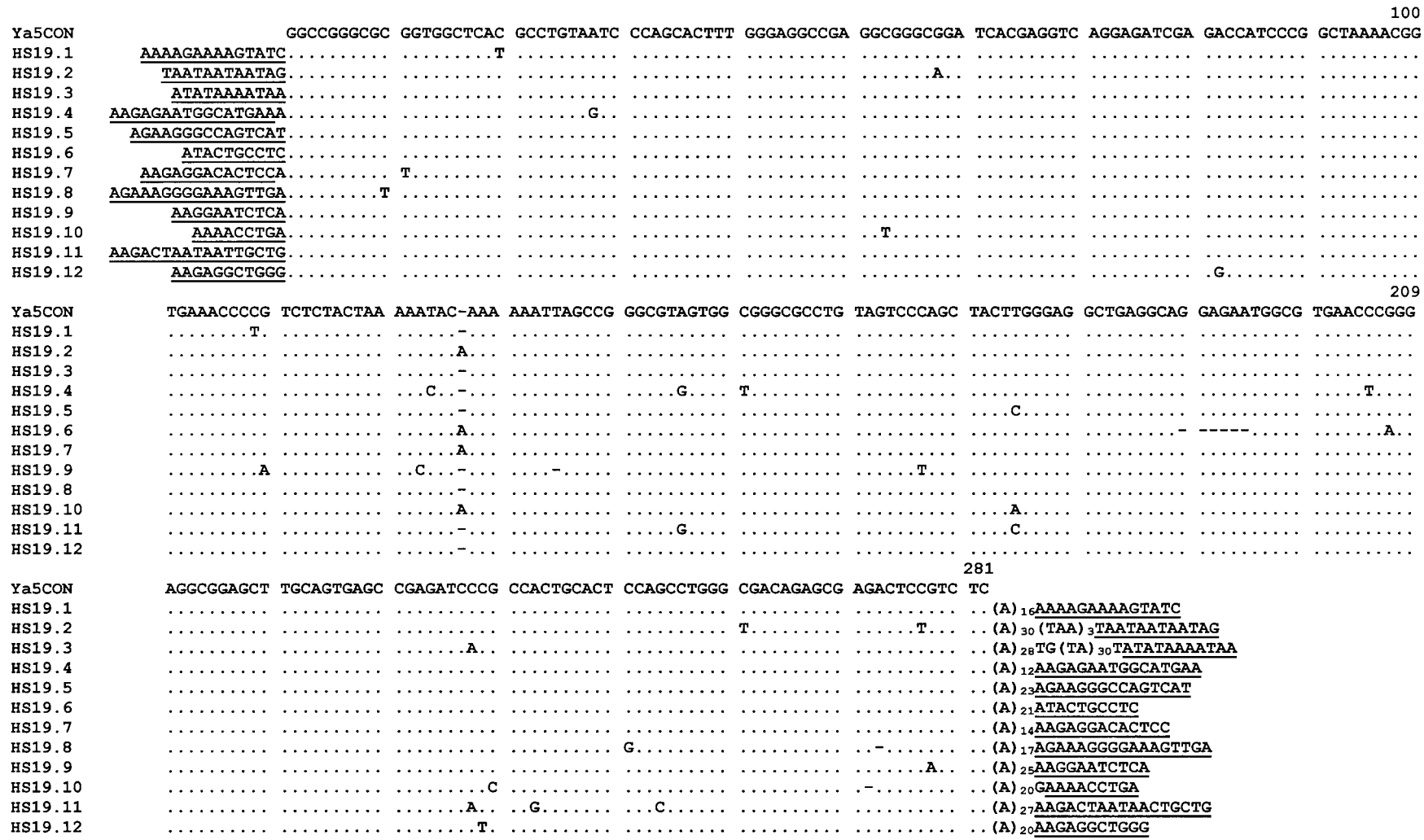

Figure 2. Alignment of chromosome 19-specific Ya5/8 Alu elements and their flanking direct repeats (underlined). The subfamily consensus sequence is shown on top (Ya5CON). Sequences that are identical to the consensus are shown as dots, substitutions are indicated by the appropriate nucleotide and deletions are shown as dashes. The composition of the 3′ oligo(dA)-rich tail is shown. HS19.9 is the only Alu element that belongs to the Ya8 subfamily while the rest belong to the Ya5 subfamily.

members as previously described (Arcot *et al.*, 1996). The CpG positions are eliminated from consideration because they mutate at a rate that is nine to ten times faster than non-CpG positions (Batzer *et al.*, 1990; Labuda & Striker, 1989) as a result of the deamination of 5-methylcytosine (Bird, 1980). Using a neutral rate of evolution of 0.15%/million years (Miyamoto *et al.*, 1987) and the level of non-CpG base substitutions of 18/3363 (0.53%) the average age of these Ya5/8 Alu repeats is estimated to be 3.56 million years old. This age estimate compares favorably with previous reports on the average age of 1.5 and 2.8 million years old since the previous estimates contained more polymorphic and presumably evolutionarily younger Alu repeats (Arcot *et al.*, 1995a and Batzer *et al.*, 1990, respectively).

Nucleotide sequences from the chromosome 19 Yb8 Alu family members are shown in Figure 3. Inspection of the nucleotide sequences shows that each of the Yb8 Alu family members was flanked by direct repeats that ranged in size from five (Sb19.9) to 20 nucleotides (Sb19.7) with an average direct repeat length of 11.9 nucleotides. In the case of Sb19.8, the direct repeat at the 5′ end has ten additional bases that appear to have inserted either after reverse transcription or subsequent to the integration of the Alu repeat. Each one of the Yb8 Alu repeats also contained an oligo(dA)-rich tail that varied from 33 (Sb19.1) to two (Sb19.6) nucleotides in length. Three of the Yb8 Alu family members (Sb19.2, Sb19.4 and Sb19.7) were truncated at the 5′ end of the Alu element but flanked by perfect direct repeat sequences, suggesting that the truncation occurred as the result of incomplete reverse transcription of the elements or an error during integration. Excluding insertions, deletions and truncations a total of 30 out of 3745 nucleotides were different from the Yb8 subfamily consensus sequence. Using the number of mutations at non-CpG dinucleotides (11/3745, 0.29%) and a neutral rate of evolution of 0.15%/million years (Miyamoto *et al.*, 1987), the average age of these Yb8 Alu repeats is estimated to be 1.93 million years old. This estimate is in agreement with previously reported age estimates of 2.7 and 4.1 million years old since some of the chromosome 19 Alu elements in this report are younger polymorphic elements (Batzer *et al.*, 1995 and Hutchinson *et al.*, 1993, respectively). Both the copy numbers and average ages of the Ya5/8 and Yb8 Alu subfamilies are similar suggesting that these two subfamilies of Alu repeats dispersed throughout the human genome at a similar rate in an overlapping evolutionary time frame.

It is interesting to note that the tails of HS19.2 and HS19.3 contain a $(TAA)_3$ and $(TA)_{30}$ simple sequence repeat, respectively. In addition, Sb19.3 and Sb19.14 contain six base insertions within their middle A-rich regions. The oligo(dA)-rich tails and middle A-rich regions of Alu elements have previously been shown to be the nuclei for the genesis of simple sequence repeats (Arcot *et al.*, 1995b; Economou *et al.*, 1990). Although we have only reported two examples, it is also worth noting that the presence of two Yb8 Alu elements with six base insertions within the middle A-rich region (Sb19.3 and Sb19.14, Figure 3) may be the first indication of a new human lineage of the Yb subfamily. This subfamily would be designated as Yb9 based upon the nine subfamily-specific mutations within these Alu elements.

Phylogenetic and genomic diversity analyses

In order to perform PCR-based analyses of the phylogentic and human genomic diversity associated with each of the chromosome 19-specific Alu family members we predicted oligonucleotide primers complementary to the unique DNA sequences adjacent to each Alu element (Table 2). With the exception of Alu elements HS19.4, Sb19.2, Sb19.8 and Sb19.11, oligonucleotide primers complementary to the rest of the Alu elements were successful in amplifying specific Alu insertion loci using genomic DNA as template for the PCR. Computer-based searches using the basic local alignment search tool (BLAST) algorithm showed that the HS19.4, Sb19.2, Sb19.8 and Sb19.11 Alu repeats had inserted within older pre-existing Alu elements. Therefore, PCR of these loci only generated specific amplicons using DNA from the original cosmid clones as templates. Amplification of these loci using genomic DNA templates produced smeared PCR products consistent with non-specific priming from multiple Alu repeats dispersed throughout the genome. The details of phylogenetic analysis on the remainder of the Alu elements using a PCR-based assay and a panel of primate genomic DNA samples showed that with the exception of HS19.8, all Alu elements were restricted in distribution to the human genome and absent from the genomes of non-human primates. HS19.8 appeared to be present in chimpanzee but was absent in gorilla and orangutan, and other primates that diverged even earlier.

The initial levels of genetic diversity of the chromosome 19 Alu elements were assessed by PCR analysis using a pop-tube approach (Ruano *et al.*, 1994) with three different pooled DNA samples from ten Caucasians, ten African-Americans and ten Hispanic-Americans. These populations sample the diversity present in Europe, Africa and the New World. Alu elements Sb19.3, Sb19.10 and Sb19.12 were classified as polymorphic using this approach. The distribution of the polymorphic Alu elements was subsequently determined using a larger number of individuals from several diverse populations (Table 3). All three polymorphic Alu elements were found in all ten populations analyzed suggesting that they integrated within the human genome prior to the

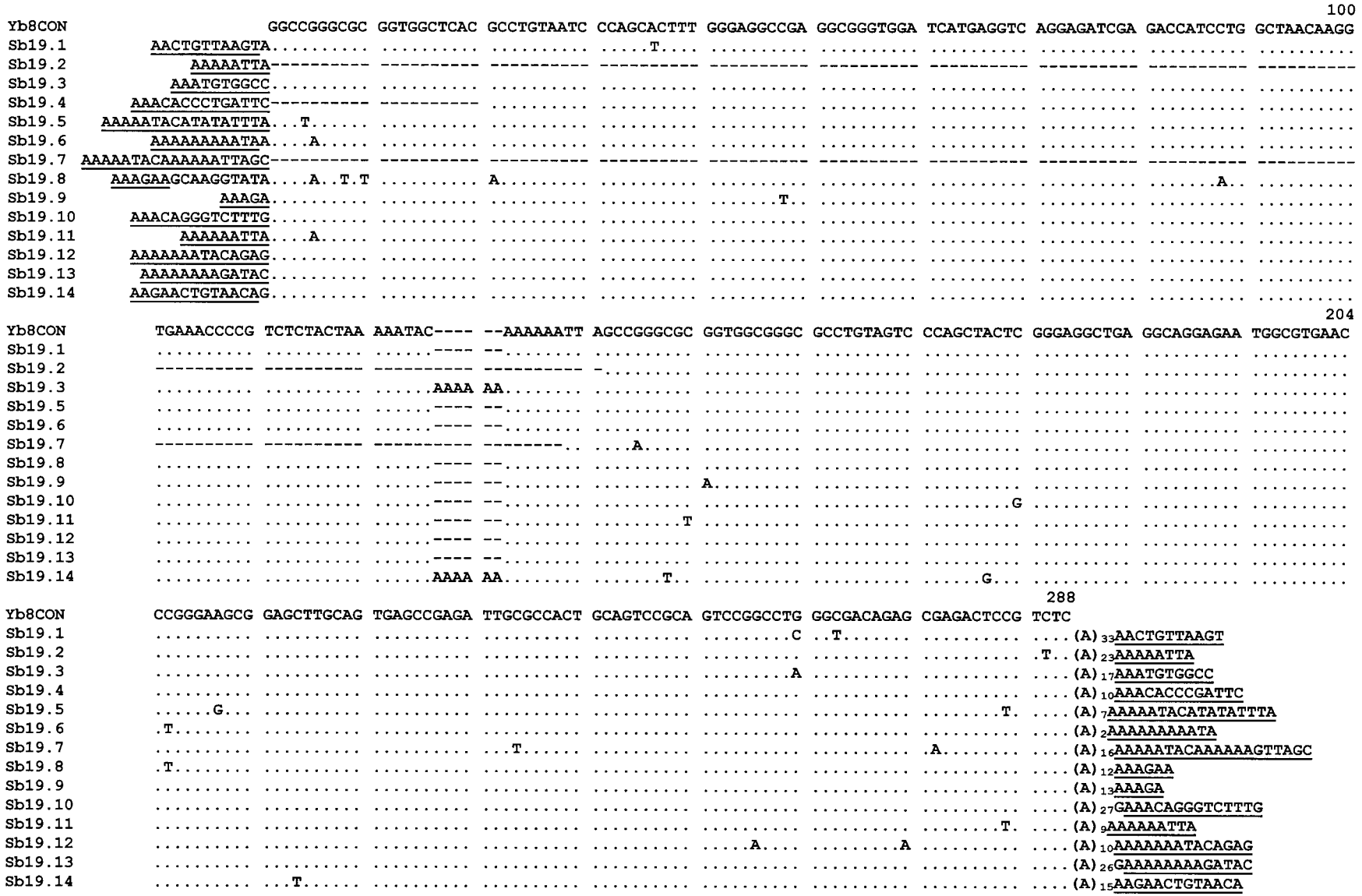

Figure 3. Alignment of chromosome 19-specific Yb8 Alu elements and their flanking direct repeats (underlined). The subfamily consensus sequence is shown on top (Yb8CON). Sequences that are identical to the consensus are shown as dots, substitutions are indicated by the appropriate nucleotide and deletions are shown as dashes. The composition of the 3′ oligo(dA)-rich tail is shown.

Table 2. Oligonucleotides for PCR amplification, annealing temperatures and polymorphism status of chromosome 19-specific young Alu repeats

Name	5′ Primer (5′-3′)	3′ Primer (5′-3′)	Annealing temp.	Polymorphism status
HS19.1	TTGTATCGAAGAGGAAGCAGGC	GCTTTACCATGATGCTGCACGGTG	57	Monomorphic
HS19.2	ACTAGAGCCCAGGAATTAGCGAC	GAATACTAGGATCTGCTTGCCTCA	65	Monomorphic
HS19.3	TTTAGCCAAGCATGGTGGCAGATG	GCTGATTTGGGAGAATGAACGACTG	67	Monomorphic
HS19.4[a]	AAAAATTAGCCAGACGTGGTGG	ACCATCAGGCTTTTCTAAGAGC	60	ND
HS19.5	AACTAAGGGCAGAGATAGCATCTAC	GTGCTAGGATTATAGGTGTGAAC	60	Monomorphic
HS19.6	CCACAATACTTATGAAGGACACC	TATTTTTCTCAGCTCGCACTACC	67	Monomorphic
HS19.7	TGTATTTCATTCACAGGGGAG	TTAACCCCAGCATTAACTTCC	63	Monomorphic
HS19.8	GAAATCTTAATGCCCAGTGCC	TGTGAGTCACCATGCTCAGCTCAAC	65	Monomorphic
HS19.10	GCCTGGCCCCATTTTCACACTGC	CTTTGTGATCCACCTGCCTTGGT	65	Monomorphic
HS19.11	ACGTAAGTGTTTTCCTCTTTG	GATCTCCAGCAGTTATTAATC	55	Monomorphic
HS19.12	AGCATCCCTGTCAACAGTGG	ACCACACCCAGCCCAGCCTCTTTTT	65	Monomorphic
Sb19.1	AACCTTCTTGGATGACAAAGCC	GCTATGATTGTGTTATTGCACTCC	60	Monomorphic
Sb19.2[a]	AATTAGCCAGGTGTGGTGGTGCAT	CAGGACACAAGCAGCGACCATT	65	ND
Sb19.3	TCTAGGCCCAGATTTATGGTAACTG	AAGCACAATTGGTTATTTTCTGAC	60	Polymorphic
Sb19.4	TCTCCAGGCTTGCACCTGGACG	TTCTCCAGGGCACTCATGTCTTCC	60	Monomorphic
Sb19.5	AATCACTTGAAGCCAGGCAACA	ACTAGTGGATCTTTCCATATGCC	60	Monomorphic
Sb19.6	GGGAGGGACAGGGAAAGATGTG	GAACTTATTCCTCCTCCTATCTAGCTG	65	Monomorphic
SB19.7	TCACCTTCCCATAACCTCTTCC	ACTGCACCTGGCCAAGATGCAT	65	Monomorphic
Sb19.8[a]	ACAGAGGGAGACTGTCTCCAG	CCTGGGCTCAAGCTATGATCC	65	ND
Sb19.9	CTGGGCAATATAGCGAGGCTC	GTCTTCCCAGCCATTCTCTGG	65	Monomorphic
Sb19.10	GCCAAGAAATAAAATAGCCCTCC	TGGGATTACAGGCATAAGCCAACAT	65	Polymorphic
Sb19.11[a]	CAGGTTCAGTATTCAGCTTGAGG	CTGCATGTTACAGTCCCTGG	60	ND
Sb19.12	TTAACATCCCTGCAACCCATC	GATTATAGTCACCCTGTTGTGC	65	Polymorphic
Sb19.13	ATGAAATTGTTTGATCCATAAATGC	GTCTTAAGAATTGCCAGCATAC	65	Monomorphic
Sb19.14	ACTCCTGAAGTCAGCAAGACCACG	CTAGGTAAGGCTACTGTGTCTCC	65	Monomorphic

[a] These primers produce the correct amplicon using cosmid DNA as a template for the PCR, but not when total genomic DNA is used as a template.

radiation of modern humans. The allele frequencies of the polymorphic elements ranged from 0.958 for Sb19.3 in U.S. Caucasians to 0.059 for Sb19.10 in Turkish Cypriots (Table 3). Each of the polymorphic Alu elements should prove useful as markers for the study of human population genetics.

Characterization of the Alu HS19.8 region

An orthologous locus that contains a young Alu element in chimpanzee and human genomes may be an active Ya5/8 source gene (Leeflang *et al*., 1992) or the product of a gene conversion of a pre-existing older Alu element to a young Alu repeat in the human genome. No gene conversion events have ever been previously reported for Ya5/8 Alu elements, however a low level of gene conversion has been observed in Yb8 Alu elements (Batzer *et al*., 1995; Kass *et al*., 1995). In order to determine the molecular nature of this locus, orthologous PCR products from human, chimpanzee, gorilla and orangutan genomes were cloned and sequenced. The alignment of the nucleotide sequences from the orthologs indicates that two Alu elements belonging to different subfamilies have independently inserted within a 50 bp region on chromosome 19 (Figure 4). The human Alu element belongs to the Ya5 subfamily while the chimpanzee Alu repeat belongs to the older Y subfamily. The insertion of the chimpanzee Y Alu element appears to have occurred independently of the Ya5 Alu insertion since each Alu repeat has unique pre-integration sequences and flanking direct repeats. In addition, the human locus has the intact pre-insertion sequence for the Y Alu element insert in chimpanzee, and *vice versa*. An inspection of the pre-integration sequences, which are 43 bases apart, does not indicate any similarity between the two sites. Therefore we conclude that the insertion of the two Alu repeats within this region are independent events. A schematic diagram of the evolution of this locus is shown in Figure 5 with the original locus devoid of an Alu element in orangutan and gorilla genomes along with the independent insertions of a Ya5 Alu repeat in humans and Y Alu family member in chimpanzees.

Discussion

Since the first identification of subfamilies of Alu elements that are active in retroposition, a number of recently integrated ''young'' Alu elements have been identified (Arcot *et al*., 1995a, 1996). Most of these elements were identified from randomly sheared genomic DNA libraries that enriched for specific subfamily members (Batzer & Deininger, 1991; Batzer *et al*., 1990). The availability of physical maps for entire human chromosomes allows one to study the dispersal of these elements from the perspective of a single chromosome and to address aspects of the biology of retroposition, such as bias towards a certain sequence context or regions of the chromosome for integration. The physical map of human chromosome 19 developed

Table 3. Distribution of the polymorphic Alu elements

Population	Genotype[a]	Sb19.3			Sb19.10			Sb19.12		
		Observed individuals	Expected[b] individuals	Allele frequencies	Observed individuals	Expected individuals	Allele frequencies	Observed individuals	Expected individuals	Allele frequencies
U. S. Caucasians	+ +	65	65.2	0.958	10	7.1	0.317	7	6.5	0.303
	+ −	6	5.7		25	30.7		29	30	
	− −	0	0.1	0.042	36	33.1	0.683	35	34.5	0.697
African Americans	+ +	18	17.5	0.507	1	0.4	0.071	12	14.6	0.457
	+ −	33	34		8	9.2		40	34.7	
	− −	17	16.5	0.493	61	60.4	0.929	18	20.6	0.543
Hispanic Americans	+ +	48	47.4	0.817	3	5.4	0.275	1	2	0.167
	+ −	20	21.2		33	28.3		22	20	
	− −	3	2.4	0.183	35	37.3	0.725	49	50	0.833
Acadians	+ +	49	47.9	0.879	3	2.1	0.179	3	1	0.120
	+ −	11	13.2		18	19.7		11	15	
	− −	2	0.9	0.121	46	45.1	0.821	57	55	0.880
Alaska Natives	+ +	28	27.9	0.755	2	1.1	0.150	0	0	0.025
	+ −	18	18.1		11	12.8		2	2	
	− −	3	2.9	0.245	37	36.1	0.850	38	38	0.975
Greenland Natives	+ +	34	31.2	0.790	3	2.2	0.210	1	0.2	0.071
	+ −	11	16.6		15	16.6		5	6.5	
	− −	5	2.2	0.210	32	31.2	0.790	43	42.3	0.929
French	+ +	61	60.5	0.917	7	6.7	0.306	7	6.7	0.318
	+ −	10	11		30	30.6		28	28.6	
	− −	1	0.5	0.083	35	34.7	0.694	31	30.7	0.682
Bretons	+ +	52	51.6	0.898	3	5.3	0.271	2	4.3	0.243
	+ −	11	11.7		33	28.4		31	26.4	
	− −	1	0.7	0.112	36	38.3	0.729	39	41.3	0.757
Turkish Cypriots	+ +	47	46.7	0.890	1	0.2	0.059	1	2.5	0.221
	+ −	11	11.6		4	5.6		21	17.9	
	− −	1	0.7	0.110	46	45.2	0.941	30	31.6	0.779
Greek Cypriots	+ +	39	37.4	0.902	5	7	0.391	3	2.4	0.238
	+ −	5	8.1		26	21.9		14	15.2	
	− −	2	0.4	0.098	15	17.1	0.609	25	24.4	0.762

[a] The presence and absence of the Alu element are indicated by + and −, respectively.
[b] Expected values are based upon the Hardy-Weinberg equilibrium.

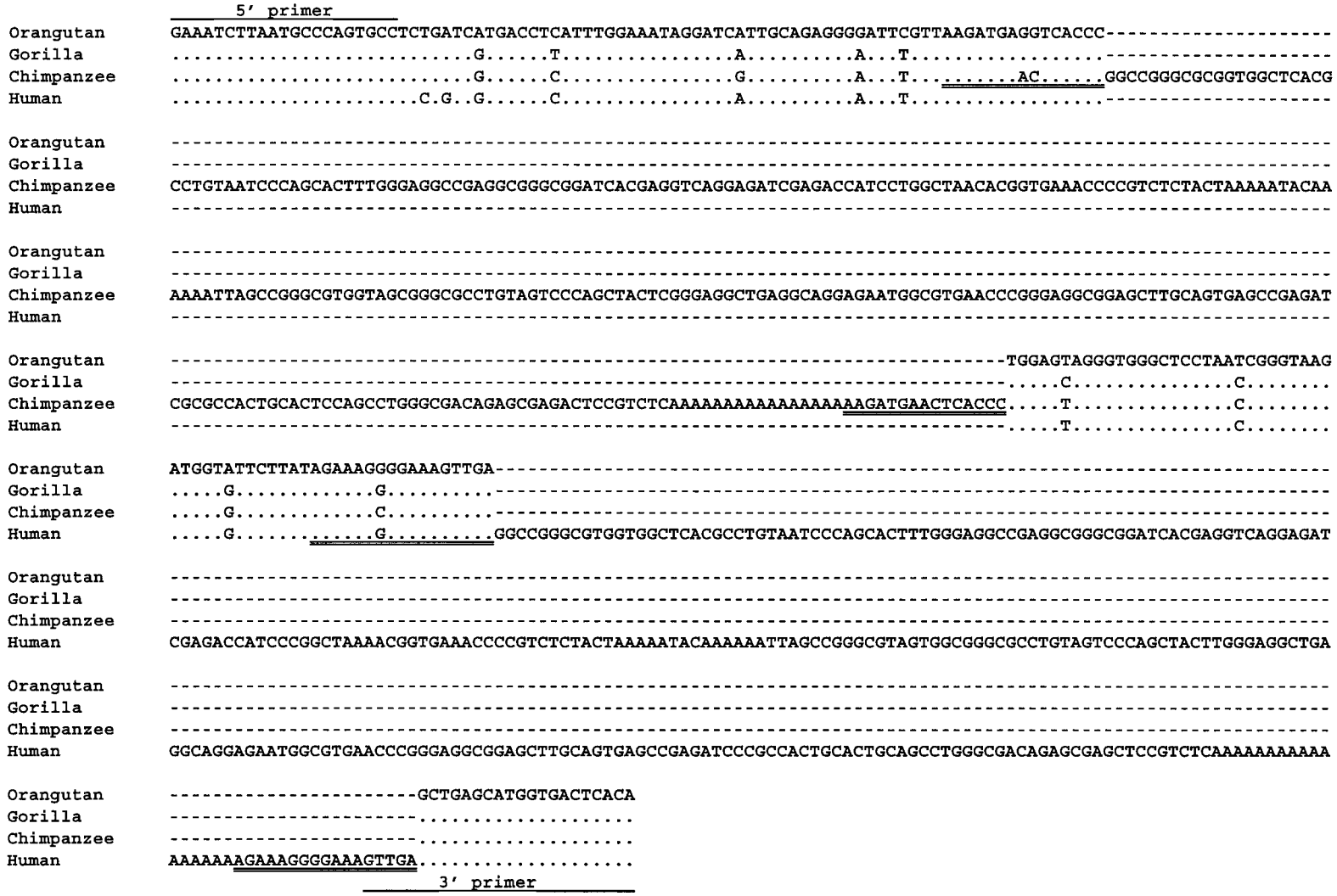

Figure 4. Alignment of the nucleotide sequences of the genomic region containing Alu element HS19.8. DNA sequences from orthologous regions of the orangutan, gorilla, chimpanzee and the human genome are shown. Dots indicate regions of identity, dashes indicate absence of sequence and nucleotide substitutions are indicated appropriately. The sequences of the primers used for generating the PCR products are shown. The sequence that is duplicated to form the flanking direct repeats upon the insertion of an Alu element is indicated with double underlines.

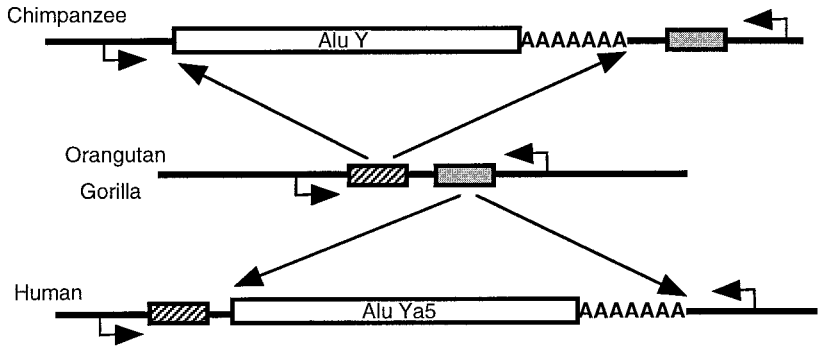

Figure 5. Schematic diagram of the evolution of Alu element HS19.8 and the adjacent DNA sequences. The pre-insertion locus corresponding to Alu element HS19.8 with the sequences that form the flanking direct repeats upon the insertion of the Alu element are represented as cross-hatched and shaded rectangular boxes in orangutan and gorilla. The independent insertion of a Y Alu element in the chimpanzee genome and a Ya5 Alu element in human genome are shown as the open rectangular boxes along with the 3′ oligo(dA)-rich tail. The bent arrows indicate the location of the oligonucleotide primer sequences that are identical in all the primates shown in the Figure.

at LLNL has a foundation primarily consisting of cosmid clones that are organized into multiple contigs (Ashworth *et al.*, 1995). These clones were arrayed onto nylon filters and screened using subfamily-specific oligonucleotide probes under stringent hybridization conditions.

Two Alu elements, PV30B and LDLR, which are known to be present on chromosome 19 (Batzer *et al.*, 1995) were not detected in this study. There are several potential reasons for not detecting some chromosome 19-specific Alu elements within the library. First, some genomic regions may be resistant to cloning or underrepresented in this particular library. The centromere and telomeric regions of chromosome 19 were underrepresented in this cosmid library since the restriction enzyme used to construct the library infrequently restricts DNA from tandem repeat sequences that are predominant in these regions of the chromosome (De Jong *et al.*, 1989). Secondly, random mutations within the region complementary to the hybridization probe may also result in the lack of identification of an Alu element. For instance, the low density lipoprotein receptor (LDLR) Alu element was not detected in this study because of the mutation in the probe hybridization site and the stringent hybridization conditions employed for the isolation of these Alu elements.

The Ya5/8 and Yb8 Alu subfamilies are composed of approximately 2000 members (Batzer *et al.*, 1995). Human chromosome 19 is very G+C-rich and also extremely gene-dense (Ashworth *et al.*, 1995). These two characteristics would suggest that Alu elements, which preferentially insert in these regions, should be present in a higher number on chromosome 19 as compared to the rest of the genome. In order to determine if there is any bias in the number of Ya5/8 and Yb8 Alu repeats located on human chromosome 19, we compared the total number of recently integrated Alu elements to that expected based upon the size of human chromosome 19. Since human chromosome 19 is approximately 2% of the human genome by mass (Ashworth *et al.*, 1995), we expect a total of 40 (2% of 1000 × two different subfamilies) recently integrated Alu family members (20 Ya5/8 and 20 Yb8) to be located along human chromosome 19. The number of observed Ya5/8 (13 total; 12 + PV30B) and Yb8 (15 total; 14 + LDLR) Alu family members does not significantly differ from that expected based upon the random integration of these Alu family members into the chromosome as judged by a chi-square test for goodness of fit. Therefore, irrespective of the gene density and the G+C content of the chromosome, we conclude that there has not been a preferential accumulation of young Alu elements on human chromosome 19 as compared to the entire genome.

In order to determine if the number of Ya5/8 and Yb8 Alu family members on each arm of chromosome 19 was significantly different from that expected, we divided the total number of elements observed from each subfamily into two classes determined by the length of each arm. Excluding the centromere and the telomeres, the p arm of chromosome 19 is approximately 20 Mb in length while the q arm is 30 Mb of the total 50 Mb. Therefore we would expect 20/50 or 40% of the recently inserted Alu repeats to be located on the p arm of the chromosome and 30/50 or 60% of the elements to be located on the q arm. A total of 27 (26 reported here + LDLR) recently integrated Alu elements were identified and mapped, therefore the p arm should contain 10.8 Alu elements (5.4 Ya5/8 and 5.4 Yb8) while the q arm should contain 16.2 Alu elements (8.1 Ya5/8 and 8.1 Yb8). An examination of Figure 1 shows that five Ya5/8 and four Yb8 Alu family members were located on the p arm of the chromosome and seven Ya5/8 and 11 Yb8 Alu elements were located on the q arm of the chromosome. The number of Alu elements from the p arm (nine) and q arm (18) of chromosome 19 does not significantly differ from that expected on each arm of the chromosome as judged by a chi-

square goodness of fit test. Therefore, we conclude that there is no preferential integration of recently integrated Ya5/8 or Yb8 Alu family members on either arm of human chromosome 19.

Previously, we have reported on the random dispersal of Ya5/8 Alu elements throughout the human genome (Arcot *et al*., 1995a, 1996). The random dispersal of recently integrated Alu repeats throughout the human genome, and along human chromosome 19 reported here provide support for the random dispersal of the young Alu elements throughout the human genome. However, Korenberg & Rykowski (1988) have previously reported an R-band specific preferential distribution for Alu repeats, and Sainz *et al*. (1992) report a non-random distribution of Alu and LINE elements from human chromosome 21. The previous study by Korenberg & Rykowski (1988) was conducted using a generic Alu repeat probe and fluorescence *in situ* hybridization to look at the composite distribution of Alu sequences within the human genome. Older Alu repeats may have had a very small bias in their distribution that results in the appearance of R-band-specific distribution using FISH with a consensus Alu probe for detection. The dispersal of Alu sequences may have favored R-bands early in primate evolution when the J and Sx Alu subfamilies amplified and the rate of Alu amplification was 100 times higher than the current rate (Shen *et al*., 1991). Subsequently, as the genome became infected with an increasing number of Alu elements, the bias toward insertion in R-bands may have changed or been lost resulting in a random pattern of integration for young Ya5/8 and Yb8 Alu repeats. The study by Sainz *et al*. (1992) was based on an analysis of *Not*I restriction fragments, and may reflect the non-random distribution of *Not*I sites along human chromosome 21 rather than just the distribution of interspersed repeats. Alternatively, the endonucleolytic activity associated with LINE elements (Feng *et al*., 1996) may have undergone mutations that altered its nucleotide recognition sequence slightly changing the integration site preference for young Alu and LINE elements over time. However, the direct repeats of Alu elements of various ages all appear to have a common recognition sequence and do not support the idea of a change in the endonucleolytic activity throughout primate evolution (Jurka, 1997).

Given the distribution of the Alu elements along the two arms of chromosome 19, there do not appear to be any hotspots for the integration of young Alu elements along the chromosome. The sole exception is the HS19.8 region, where an older Y Alu element has independently inserted in the same genomic region in the chimpanzee genome. This chimpanzee Y Alu insertion appears to have occurred in recent evolutionary history since it is absent in gorillas and orangutans. Considering the size of the human genome (3×10^9 bp), the number of Ya5/8 (1000) and Y (75,000) Alu repeats, and their respective amplification rates (200/million years and 5000/million years), the probability of two independent Alu insertions occurring in the same genomic region is extremely low. Several recent Y subfamily Alu insertions in chimpanzees, gorillas and humans have been identified (Ryan & Dugaizyk, 1989; Traubuchet *et al*., 1987; Muratani *et al*., 1991) demonstrating that this lineage of Alu repeats has mobilized in recent primate evolution along with Ya5/8 and Yb8 Alu family members.

Previous studies have demonstrated that approximately 25% of all recently inserted Alu elements are polymorphic within different human population groups (Arcot *et al*., 1995a). Therefore the identification of only three polymorphic Alu elements out of 28 that are located on chromosome 19 appears to be lower than the seven expected. However, the chromosome 19-specific cosmid library was prepared from flow-sorted chromosomes 19 contained in a human-rodent somatic cell hybrid cell line. This cell line was generated using cells from a single individual and contains only a single chromosome 19 and this conditions the probability for detecting young Alu repeats to a haploid genetic background. In addition, four of the recently integrated Alu elements had older Alu elements in their flanking sequences and could not be tested for polymorphism. Nonetheless, the three polymorphic Alu elements that were identified should prove to be useful markers to study human genomic diversity.

In summary, we have performed an analysis of the recent Alu insertions belonging to two of the youngest subfamilies, on a single human autosome. The results of this study indicate that the number of elements identified on chromosome 19 compares favorably with the original estimates of the total number of Alu elements from these two subfamilies that are present in the genome. There does not appear to be a bias for insertion of young Alu elements on human chromosome 19 as a whole or into a particular arm of the chromosome. Earlier studies showed that Alu insertions have a biased distribution (Sainz *et al*., 1992) and integrate preferentially into G+C rich regions, which are generally associated with R-bands (Korenberg & Rykowski, 1988). Our results show that young Alu repeats integrate in both G and R chromosomal bands, supporting a model of random dispersal of young Alu elements throughout the genome. These data also indicate that chromosomal G+C content may not be the major factor influencing the dispersal and integration of these elements.

Materials and Methods

Cell lines and DNA samples

The cell lines used to isolate human DNA samples were as follows: human (*Homo sapiens*), HeLa (ATCC CCL2); chimpanzee (*Pan troglodytes*), Wes (ATCC CRL1609); gorilla (*Gorilla gorilla*), Ggo-1 (primary gorilla fibroblasts) provided by Dr Stephen J. O'Brien, National

Cancer Institute, Frederick, MD, USA; African Green Monkey (*Cercopithecus aethiops*), CV1 (ATCC CCL70); and owl monkey (*Aotus trivirgatus*), OMK (637-69 ATCC CRL1556). Cell lines were maintained as directed by the source and DNA isolations were performed by standard methods as described (Ausabel *et al*., 1987). Additional DNA samples from five individual chimpanzees, one gorilla, three orangutans (*Pongo pygmaeus*), one macaque (*Macaca fascicularis*) and one tamarin (*Saguinus oedipus*) were obtained from BIOS Laboratories, New Haven, CT, USA. Human DNA samples from the various population groups were isolated from peripheral blood lymphocytes (Ausabel *et al*., 1987) and were available from previous studies (Batzer *et al*., 1996a).

Library construction, screening and subcloning

The chromosome 19-specific cosmid library was generated from flow-sorted human chromosomes isolated from a Chinese hamster-human hybrid cell line (UV5HL9-5B) that contains a single human chromosome 19. Construction of the chromosome-specific library in a cosmid vector and its characterization has been described elsewhere (De Jong *et al*., 1989). Approximately 10,000 clones from this library (which represents a sevenfold coverage of the chromosome) were arrayed in duplicate onto nylon membranes (Hybond-N, Amersham, Arlington Heights, IL, USA) as described (Olsen *et al*., 1993). These clones were screened using oligonucleotide probes specific to the Ya5/8 and Yb8 subfamilies of Alu repeats (Batzer & Deininger, 1991; Batzer *et al*., 1995) using hybridization and wash conditions described by Batzer *et al*. (1995).

Subcloning of the cosmid clones containing the Alu elements to identify the smallest fragment containing the specific Alu repeat was performed using the following strategy. Plasmid pBluescript KS+ (Stratagene, La Jolla, CA, USA) was linearized with *Bam*HI and dephosphorylated. The cosmid clone was digested with a combination of restriction enzymes *Bam*HI, *Bgl*II and *Bcl*I. All three enzymes produce cohesive ends compatible for cloning into the *Bam*HI site of the vector. The mixture of fragments generated in this fashion was purified and ligated to the linearized vector, transformed into competent DH5α and plated onto 132 mm 2 × YT agar plates. Colony lifts were performed using standard protocols (Maniatis *et al*., 1989) and were screened using radiolabeled oligonucleotide probes specific for Ya5 and Yb8 Alu elements. After a second round of plating and screening, the positive clones were processed to generate templates for sequence analysis.

DNA sequencing

Plasmid templates for sequencing were prepared by Qiagen mini-alkaline lysis according to the manufacturer's instructions (Qiagen, Chatsworth, CA, USA). Double-stranded plasmid templates corresponding to the cosmid subclones containing the Alu elements were sequenced using internal Alu-specific primers as described (Batzer & Deininger, 1991; Batzer *et al*., 1995) and fluorescently labeled dye terminators using AB *Taq* cycle sequencing kits (Applied Biosystems Division, Foster City, CA, USA). Sequencing of the PCR products corresponding to the Alu elements and their flanking regions cloned into pCR II TA-cloning vector (Invitrogen, San Diego, CA, USA) was accomplished using fluorescently labeled M13 forward and reverse primers and AB *Taq* cycle sequencing kits. Sequencing reactions were fractionated on a 6% (w/v) polyacrylamide gel, followed by data collection and analysis on an Applied Biosystems 373A DNA sequencer. The GenBank accession numbers for the sequences in this report are shown in Table 1 and also include AF077056-AF077058 for the orangutan, gorilla and chimpanzee HS19.8 orthologous regions (respectively).

Primer design and PCR amplification

Sequences flanking individual Alu elements were screened against the GenBank non-redundant database for the presence of repetitive elements using the basic local alignment sequence tool (BLAST). PCR primers were designed either manually or using the software PRIMER (Whitehead Institute for Biomedical Research, Cambridge, MA, USA). PCR amplification was carried out in 50 µl reactions as previously described by Arcot *et al*. (1996). The sequences of the primers and their annealing temperatures are shown in Table 2.

Acknowledgments

We thank Julie Thiede for providing the cosmid clones, Stephanie Johnson for restriction mapping the cosmid clones and, Mari Christensen and Laurie Gordon for assistance with the FISH mapping of cosmids. This work was supported by grants from U. S. Department of Energy LDRD 94-LW-103 (M.A.B.) and National Science Foundation SBR-9610147 (M.A.B.). Work at Lawrence Livermore National Laboratory was conducted under the auspices of U. S. Department of Energy contract number W-7405-ENG-48.

References

Arcot, S. S., Shaikh, T. H., Kim, J., Bennett, L., Alegria-Hartman, M., Nelson, D. O., Deininger, P. L. & Batzer, M. A. (1995a). Sequence diversity and chromosomal distribution of ''young'' Alu repeats. *Gene,* **163**, 273–278.

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

Arcot, S. S., Adamson, A. W., Lamerdin, J. E., Kanagy, B., Deininger, P. L., Carrano, A. V. & Batzer, M. A. (1996). Alu fossil relics – Distribution and insertion polymorphism. *Genome Res.* **6**, 1084–1092.

Ashworth, L. K., Batzer, M. A., Brandriff, B., Branscomb, E., De Jong, P., Garcia, E., Garnes, J. A., Gordon, L. A., Lamerdin, J. E., Lennon, G., Mohrenweiser, H., Olsen, A. S., Slezak, T. & Carrano, A. V. (1995). An integrated metric physical map of human chromosome 19. *Nature Genet.* **11**, 422–427.

Ausabel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1987). *Current Protocols in Molecular Biology*, John Wiley & Sons, New York.

Batzer, M. A. & Deininger, P. L. (1991). A human-specific Subfamily of Alu sequences. *Genomics,* **9**, 481–487.

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

Batzer, M. A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D. H., Shaikh, T. H., Novick, G. E., Ioannou, P. A., Scheer, W. D., Herrera, R. J. & Deininger, P. L. (1994). African origin of human-specific polymorphic Alu insertions. *Proc. Natl Acad. Sci. USA,* **91**, 12288–12292.

Batzer, M. A., Rubin, C. M., Hellmann-Blumberg, U., Alegria-Hartman, M., Leeflang, E. P., Stern, J. D., Bazan, H. A., Shaikh, T. H., Deininger, P. L. & Schmid, C. W. (1995). Dispersion and insertion polymorphism in two small subfamilies of recently amplified Alu repeats. *J. Mol. Biol.* **247**, 418–427.

Batzer, M. A., Arcot, S. S., Phinney, J. W., Alegria-Hartman, M., Kass, D. H., Milligan, S. M., Kimpton, C., Gill, P., Hochmeister, M., Ioannou, P. A., Herrera, R. J., Boudreau, D. A., Scheer, W. D., Keats, B. J. B., Deininger, P. L. & Stoneking, M. (1996a). Genetic variation of recent Alu insertions in human populations. *J. Mol. Evol.* **42**, 22–29.

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

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

Brandriff, B. F., Gordon, L. A., Fertitta, A., Olsen, A. S., Christensen, M., Ashworth, L. K., Nelson, D. O., Carrano, A. V. & Mohrenweiser, H. W. (1994). Human chromosome 19p: a fluorescence *in situ* hybridization map with genomic distance estimates for 79 intervals spanning 20 Mb. *Genomics,* **23**, 582–591.

De Jong, P. J., Yokabata, K., Chen, C., Lohman, F., Pederson, L., McNinch, J. & Van Dilla, M. (1989). Human chromosome-specific partial digest libraries in lambda and cosmid vectors. *Cytogenet. Cell Genet.* **51**, 985.

Deininger, P. L. (1989). SINEs, short interspersed repeated DNA elements in higher eukaryotes. In *Mobile DNA* (Berg, D. E. & Howe, M. M., eds), pp. 619–636, American Society of Microbiology, Washington, DC.

Deininger, P. L. & Batzer, M. A. (1993). Evolution of retroposons. In *Evolutionary Biology* (Hecht, M. K., MacIntyre, R. J. & Clegg, M. T., eds), vol. 27, pp. 157–196, Plenum Press, New York.

Deininger, P. L. & Batzer, M. A. (1995). SINE master genes and population biology. In *The Impact of Short Interspersed Elements (SINEs) on the Host Genome* (Maraia, R. J., ed.), pp. 43–60, Landes, Georgetown.

Economou, E. P., Bergen, A. W., Warren, A. C. & Antonarakis, S. E. (1990). The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc. Natl Acad. Sci. USA,* **87**, 2951–2954.

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

Gordon, L. A., Bergmann, A., Christensen, M., Danganan, L., Lee, D. A., Ashworth, L. K., Nelson, D. O., Olsen, A. S., Mohrenweiser, H. W., Carrano, A. V. & Brandriff, B. F. (1995). A 30-Mb metric fluorescence *in situ* hybridization map of human chromosome 19q. *Genomics,* **30**, 187–192.

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

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

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.

Korenberg, J. R. & Rykowski, M. C. (1988). Human genome organization: Alu, LINES and the molecular structure of metaphase chromosome bands. *Cell,* **53**, 391–400.

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

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.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1989). *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

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.

Muratani, K., Hada, T., Yamamoto, Y., Kaneko, T., Shigeto, Y., Ohue, T., Furuyama, J. & Higashino, K. (1991). Inactivation of the cholinesterase gene by Alu insertion: possible mechanism for human gene transposition. *Proc. Natl Acad. Sci. USA,* **88**, 11315–11319.

Okada, N. (1991). SINEs. *Curr. Opin. Genet. Dev.* **1**, 498–504.

Olsen, A. S., Combs, J., Garcia, E., Elliot, J., Amemiya, C., De Jong, P. & Threadgill, G. (1993). Automated production of high density cosmid and YAC colony filters using a robotic workstation. *BioTechniques,* **14**, 116–123.

Ruano, G., Deinard, A. S., Tishkoff, S. & Kidd, K. K. (1994). Detection of DNA sequence variation via deliberate heteroduplex formation from genomic DNAs amplified en masse in ''population tubes''. *PCR Methods Appl.* **3**, 225–231.

Ryan, S. C. & Dugaiczyk, A. (1989). Newly arisen DNA repeats in the primate phylogeny. *Proc. Natl Acad. Sci. USA,* **86**, 9360–9364.

Sainz, J., Pevny, L., Wu, Y., Cantor, C. R. & Smith, C. L. (1992). Distribution of interspersed repeats (Alu and Kpn) on Not I restriction fragments of human chromosome 21. *Proc. Natl Acad. Sci. USA,* **89**, 1080–1084.

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

Shen, M. R., Batzer, M. A. & Deininger, P. L. (1991). Evolution of the ''master'' Alu gene(s). *J. Mol. Evol.* **33**, 311–320.

Stoneking, M., Fontius, J. J., Clifford, S. L., Soodyall, H., Arcot, S. S., Saha, N., Jenkins, T., Tahir, M. A., Deininger, P. L. & Batzer, M. A. (1997). Alu insertion polymorphisms and human evolution: Evidence for a larger population size in Africa. *Genome Res.* **7**, 1061–1071.

Traubuchet, G., Chebloume, Y., Savatier, P., Lachuer, J., Faure, C., Verdier, G. & Nigon, V. M. (1987). Recent insertion of an Alu sequence in the beta-globin gene cluster of the gorilla. *J. Mol. Evol.* **25**, 288–291.

Vidaud, D., Vidaud, M., Bahnak, B. R., Siguret, V., Sanchez, S. G., Laurian, Y., Meyer, D., Goossens, M.

& Lavergne, J. M. (1993). Haemophilia B due to a de novo insertion of human-specific Alu subfamily member within the coding region of the factor IX gene. *Eur. J. Hum. Genet.* **1**, 30–36.

Wallace, M. R., Andersen, L. B., Saulino, A. M., Gregory, P. E., Glover, T. W. & Collins, F. S. (1991). A de novo Alu insertion results in neurofibromatosis type I. *Nature,* **353**, 864–866.

Edited by J. Karn

(Received 31 March 1998; received in revised form 29 May 1998; accepted 12 June 1998)