

8-1-2003

Following the LINEs: An analysis of primate genomic variation at human-specific LINE-1 insertion sites

Bethaney J. Vincent

Jeremy S. Myers

Huei Jin Ho

Gail E. Kilroy

Jerilyn A. Walker

See next page for additional authors

Follow this and additional works at: https://repository.lsu.edu/biosci_pubs

Recommended Citation

Vincent, B., Myers, J., Ho, H., Kilroy, G., Walker, J., Watkins, W., Jorde, L., & Batzer, M. (2003). Following the LINEs: An analysis of primate genomic variation at human-specific LINE-1 insertion sites. *Molecular Biology and Evolution*, 20 (8), 1338-1348. <https://doi.org/10.1093/molbev/msg146>

This Article is brought to you for free and open access by the Department of Biological Sciences at LSU Scholarly Repository. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Scholarly Repository. For more information, please contact ir@lsu.edu.

Authors

Bethaney J. Vincent, Jeremy S. Myers, Huei Jin Ho, Gail E. Kilroy, Jerilyn A. Walker, W. Scott Watkins, Lynn B. Jorde, and Mark A. Batzer

Following the LINES: An Analysis of Primate Genomic Variation at Human-Specific LINE-1 Insertion Sites

Bethaney J. Vincent,* Jeremy S. Myers,* Hwei Jin Ho,* Gail E. Kilroy,* Jerilyn A. Walker,* W. Scott Watkins,† Lynn B. Jorde,† and Mark A. Batzer*

*Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University; and

†Department of Human Genetics, University of Utah Health Sciences Center

The L1 Ta subfamily of long interspersed elements (LINES) consists exclusively of human-specific L1 elements. Polymerase chain reaction–based screening in nonhuman primate genomes of the orthologous sites for 249 human L1 Ta elements resulted in the recovery of various types of sequence variants for approximately 12% of these loci. Sequence analysis was employed to capture the nature of the observed variation and to determine the levels of gene conversion and insertion site homoplasmy associated with LINE elements. Half of the orthologous loci differed from the predicted sizes due to localized sequence variants that occurred as a result of common mutational processes in ancestral sequences, often including regions containing simple sequence repeats. Additional sequence variation included genomic deletions that occurred upon L1 insertion, as well as successive mobile element insertions that accumulated within a single locus over evolutionary time. Parallel independent mobile element insertions at orthologous loci in distinct species may introduce homoplasmy into retroelement-based phylogenetic and population genetic data. We estimate the overall frequency of parallel independent insertion events at L1 insertion sites in seven different primate species to be very low (0.52%). In addition, no cases of insertion site homoplasmy involved the integration of a second L1 element at any of the loci, but rather largely involved secondary insertions of Alu elements. No independent mobile element insertion events were found at orthologous loci in the human and chimpanzee genomes. Therefore, L1 insertion polymorphisms appear to be essentially homoplasmy free characters well suited for the study of population genetics and phylogenetic relationships within closely related species.

Introduction

The L1 family of long interspersed elements (LINES) (Sheen et al. 2000) has amplified in mammalian lineages over the past 100 to 150 Myr of evolution (Skowronski and Singer 1986; Fanning and Singer 1987; Smit et al. 1995). These autonomous retrotransposons, which account for approximately 17% of the human genome, have directly influenced the expansion and landscape of the genome through a “copy and paste” mechanism of amplification (Smit 1999; Prak and Kazazian 2000; International Human Genome Sequencing Consortium 2001). Full-length retrotransposition competent L1 elements are about 6 kb in length, and they encode two open reading frames (ORFs). The first ORF encodes an RNA binding protein. The second ORF encodes a protein possessing both endonuclease (EN) and reverse transcriptase (RT) domains (Feng et al. 1996). The ORFs are preceded by a 5′ untranslated region (UTR), which contains an internal polymerase II promoter, and are followed by a 3′ UTR (Singer 1982; Skowronski, Fanning, and Singer 1988; Singer et al. 1993; Kazazian and Moran 1998; Malik and Eickbush 1999; Kazazian 2000). The protein products encoded by ORF 1 and ORF 2 possess a *cis* preference, which likely facilitates formation of a ribonucleoprotein complex that enables retrotransposition to occur (Hohjoh and Singer 1996; Wei et al. 2001). The primary mode of retrotransposition, termed target primed reverse transcription (TPRT), is dependent upon the EN and RT activities of the ORF 2 protein (Luan et al. 1993). The mobilization and integration of a new L1 insertion by TPRT results in the creation of

short direct sequence repeats at each end of the newly integrated L1 element, known as target site duplications (TSDs) (Luan et al. 1993; Jurka 1997; Cost and Boeke 1998).

L1 elements have had significant effects upon the overall architecture of the human genome, such as expanding the size of the genome (Salem et al. 2003), altering gene expression, disrupting coding sequences and splice sites, and providing areas of sequence identity for both gene conversion and recombination events (Kazazian et al. 1988; Yang et al. 1998; Rothbarth et al. 2001). In addition, L1 elements have demonstrated an ability to shuffle regions of the genome by a process termed three-prime transduction (Boeke and Pickeral 1999; Moran, DeBerardinis, and Kazazian 1999; Goodier, Ostertag, and Kazazian 2000; Kazazian and Cotton 2001). L1 elements have also been implicated in DNA repair processes and in overall genomic stability, as they have demonstrated an ability to repair double-strand breaks in DNA by an endonuclease-independent insertion mechanism (Morrish et al. 2002). The TPRT-based mobilization and subsequent insertion of L1 elements have also been shown to generate genomic variation through the creation of deletions and duplications (Gilbert, Lutz-Prigge, and Moran 2002).

A limited number of L1 elements have been active during primate evolution and are believed to have amplified according to a model known as the “master gene model” (Deininger et al. 1992). According to this model, L1 elements were created from a series of “source genes” that gradually accumulated diagnostic base changes. These changes and subsequent amplification from source genes have resulted in a series of L1 sequences that contain shared sequence variants that define them as a subfamily. In addition, because of the time frame over which these changes occur within source genes, individual L1 subfamilies appear to be of different genetic

Key words: mobile elements, parallel insertions, homoplasmy free.

E-mail: mbatzer@lsu.edu.

Mol. Biol. Evol. 20(8):1338–1348. 2003

DOI: 10.1093/molbev/msg146

Molecular Biology and Evolution, Vol. 20, No. 8,

© Society for Molecular Biology and Evolution 2003; all rights reserved.

ages (Deininger et al. 1992; Batzer, Schmid, and Deininger 1993; Deininger 1993; Smit et al. 1995). Due to their mode of amplification, L1 elements have created interspecies as well as intraspecies genetic differences. L1 insertions have properties similar to other mobile element insertions because they are known to be relatively stable upon insertion, have known ancestral states, and are thought to be identical by descent (IBD) characters for the study of population genetics and phylogenetic relationships (Sheen et al. 2000; Myers et al. 2002; Ovchinnikov, Rubin, and Swergold 2002; Roy-Engel et al. 2002; Mathews et al. 2003; Salem et al. 2003).

The most recently integrated subfamilies of human L1 elements were identified as a result of their sequence identity to known retrotransposition competent elements and disease-causing de novo inserts (Kazazian et al. 1988; Woods-Samuels et al. 1989; Schwahn et al. 1998; Meischl et al. 2000). We have previously completed a detailed study of the youngest known subfamilies of L1 elements, termed Ta (Transcribed, subset a) (Myers et al. 2002) and preTa (Salem et al. 2003). Collectively these L1 subfamilies are composed of several hundred elements, with many representing insertion polymorphisms in diverse human genomes (Boissinot, Chevret, and Furano 2000; Myers et al. 2002; Salem et al. 2003). Mobile element insertions are thought to be largely homoplasmy free characters aside from the rare independent parallel insertion of two mobile elements into identical target sites in multiple genomes (Cantrell et al. 2001; Roy-Engel et al. 2002). However, the homoplasmy free nature of mobile element insertion polymorphisms has been questioned (Hillis 1999). Here, we report an examination of 254 orthologous L1 Ta insertion sites in nonhuman primates to directly determine the levels of gene conversion and insertion site homoplasmy associated with LINE elements.

Materials and Methods

DNA Samples and Cell Lines

Primate DNA samples were obtained from the Coriell Cell repositories as follows: common chimpanzee (*Pan troglodytes*), repository number NG06939; pygmy chimpanzee (*Pan paniscus*), repository number NG05253; lowland gorilla (*Gorilla gorilla*), repository number NG05251; orangutan (*Pongo pygmaeus*), repository number NG12256; rhesus macaque (*Macaca mulatta*), repository number NG07109; pig-tailed macaque (*Macaca nemestrina*), repository number NG08452; red-bellied tamarin (*Saquinus labiatus*), repository number NG05308; woolly monkey (*Lagothrix lagotricha*), repository number NG05356; black-handed spider monkey (*Ateles geoffroyi*), repository number NG05352; and ring-tailed lemur (*Lemur catta*), repository number NG07099. Additional cell lines were maintained as directed by the source and used to isolate DNA from the following primate species: green monkey (*Cercopithecus aethiops*), ATCC number CCL70; owl monkey (*Aotus trivirgatus*), ATCC number CRL1556; and galago (adenovirus 12 SV40-transformed *Galago senegalensis* fibroblasts). Human DNA samples, used in prior studies, were isolated from peripheral blood lymphocytes (Ausubel et al. 1987; Myers et al. 2002).

PCR Design and Analysis

PCR primer design and analysis were performed as previously described (Myers et al. 2002). A complete list of L1 Ta loci examined in this study, along with primer sequences, and annealing temperatures can be found as online Supplementary Material at the journal's Web site and at <http://batzerlab.lsu.edu>.

PCR amplification of the L1HS337 locus was accomplished using the Roche Expand Long Template PCR. These consisted of 50 μ l reactions prepared according to the manufacturer's instructions, including 200 to 500 ng of template DNA and 0.3 μ M of each primer as listed previously (Myers et al. 2002). The reaction cycle consisted of an initial denaturation period at 94°C for 120 s, 30 cycles of 10 s denaturation at 94°C, 30 s annealing at 60°C, and 180 s elongation at 68°C, followed by 10 min final extension at 68°C. For analysis, 10 μ l of each sample and 10 μ l of RediLoad (Research Genetics) were fractionated on a 1% agarose gel with 0.05 μ g/ml ethidium bromide. This method was also used in an attempt to amplify preintegration sites for the following elements: L1HS327, L1HS405, L1HS469, L1HS498, and L1HS564.

Cloning and Sequence Analysis

Gel purified L1-related PCR products were cloned using Invitrogen's TOPO TA Cloning Kit according to the manufacturer's instructions. DNA sequence analysis of the cloned PCR products was accomplished by the chain-termination method using an automated ABI Prism 3100 sequencer (Sanger, Nicklen, and Coulson 1977). One hundred thirteen DNA sequences were assigned GenBank accession numbers AY246432 to AY246544. The sequences AF461364 to AF461389 from previous studies were also used in this analysis (Morrish et al. 2002). Sequence alignments were generated using MegAlign software (DNASar version 5.0) and can be found as online Supplementary Material and at <http://batzerlab.lsu.edu>.

Computational Analysis

L1HS Ta loci that experienced parallel insertion events in other nonhuman primate genomes (L1HS45, L1HS363, L1HS413, L1HS480, L1HS521, and L1HS561) were analyzed and annotated for sequence composition and proximity to coding sequence using BLAT searches (UCSC Genome Bioinformatics) of the human genome "November 2002 assembly of sequence," using the default parameters. In addition, 25,000 bp of sequence flanking the L1 Ta insertions were also analyzed for GC content using EditSeq software (DNASar version 5.0).

Results

Genomic Deletions

Two hundred fifty-four loci that were confirmed to contain an L1 Ta element in the human genome were analyzed by orthologous PCR analysis in a diverse array of nonhuman primates. Collectively, these species capture 65

Myr of primate evolution, including representatives of African apes, Old World monkeys, New World monkeys, and prosimians. Preintegration sites, whose sequences are reconstructed computationally by removing the L1 element and one of its associated TSDs (Myers et al. 2002), were amplified for 249 of the 254 L1 Ta loci. When the initial PCR analysis generated amplicons that differed in size from that predicted by sequence derived from the human genome, additional nonhuman primate species were assayed. These additional species included Old World monkeys (rhesus macaque and pig-tailed macaque), New World monkeys (woolly monkey, black-handed spider monkey, and red-bellied tamarin), and another prosimian, the ring-tailed lemur.

The preintegration sites for five elements (L1HS327, L1HS405, L1HS469, L1HS498, and L1HS564) did not amplify in any nonhuman primate species. Previously, the insertion of L1 elements has been shown to be associated with large genomic deletions (Gilbert, Lutz-Prigge, and Moran 2002). Thus, one possible explanation for the absence of preintegration PCR products would be that a large deletion (>1 kb) occurred at each of these loci during L1 integration. If a deletion occurred during the integration of the L1 elements in the human genome, then the preintegration product sizes calculated computationally would be underestimates of the true size of each locus. To investigate this possibility, we utilized long template PCR reactions of these loci that would facilitate the amplification of larger (up to 25 kb) products. Unfortunately, PCR amplicons were not generated by any of these loci, suggesting that the retrotransposition of these L1 elements in humans may have generated deletions greater than 25 kb in size. Alternately, the orthologous loci in nonhuman primate genomes may have undergone sequence changes at the primer sites, preventing PCR amplification.

We have also isolated five smaller potential L1-mediated deletion events in the human genome, including L1HS337 (> 300 bp), L1HS178 (3 bp), L1HS242 (8 bp), L1HS443 (1 bp), and L1HS513 (1 bp). The two definitive examples of very small deletions (L1HS178 and L1HS242) resulting from the L1 insertion event identified in this study were discovered through sequence analysis that was performed due to another suspected variation within the locus in nonhuman primates and/or in humans possessing an empty site. There are likely many more instances of 1-bp to 10-bp deletions that result in shifts between observed and expected preintegration product sizes that are too small to detect upon gel electrophoresis and UV visualization.

The L1HS337 locus (GenBank accession numbers AY246490 to AY246494) has a deletion of approximately 375 bp that occurred when the Ta L1 element integrated at this locus (fig. 1A and B). PCR analysis of the locus in humans and nonhuman primates generated a preintegration site amplicon in humans and in nonhuman primates that was about 375 bp larger than the predicted product size. In addition, the endonuclease cleavage site for this locus as derived from the GenBank sequence appeared to be an atypical TTTA/T, which may also be indicative of an L1 locus involved in a genomic deletion. Sequence analysis confirmed the deletion of 372 bp of unique, nonrepetitive

sequence that is found at the integration site in all nonhuman primates as well as in the empty human allele.

Short Sequence Variants

Several nonhuman primate preintegration sites differed in size from those predicted based upon the draft human genomic sequence. The first type of alteration consists of 12 different small variations spanning less than 100 bp in length that occur within the preintegration sites of one or more primate species (table 1). Most of these variants were confined to a single primate species or were shared by closely related species. In addition, the majority of these localized sequence variants were present in loci of primates that diverged from humans over 25 MYA and would have had a total of 50 Myr of time to accumulate new mutations (Goodman 1999). Sequence analysis resulted in the recovery of three additional L1 Ta preintegration loci (L1HS284, L1HS348, and L1HS411) that appear to be undergoing dynamic variation throughout the primate lineages (table 2). These loci contain regions of microsatellites or other low complexity sequences that expand and contract over evolutionary time, likely as a result of replication slippage and/or crossover events (Levinson and Gutman 1987; Schlotterer and Tautz 1992).

Successive Mobile Element Insertions

Five primate L1 Ta preintegration loci (L1HS86, L1HS135, L1HS155, L1HS301, and L1HS361) were variable due to the presence or absence of other types of mobile elements that integrated within the oligonucleotide primer sites for the L1 Ta locus of interest. In some cases, these neighboring mobile element insertions are relatively young L1 and Alu elements that are not found in nonhuman primate species that diverged before the amplification of these particular mobile element subfamilies. Therefore, sequence analysis of these loci allows us to track the successive integration of retroelements into a limited genomic region over the course of primate evolution (table 3). It is important, however, to distinguish these events from true independent parallel insertions, which will be enumerated in a later section, since retroelements themselves may provide a preferable target site for the subsequent integration of other retroelements (Salem et al. 2003). This is particularly true when the original mobile element is an L1-related element as a result of its A/T-rich sequence that provides a preferable target site, as do the A-rich tails and middle A-rich regions of preexisting genomic Alu elements (Jurka 1997).

The L1HS301 locus (GenBank accession numbers AY246482 and AY246485) allows us to reconstruct the insertion history of multiple mobile elements into a small genomic region from before the origin of the primate order up until the past few million years in which a polymorphic L1HS Ta element inserted into the human lineage (figs. 2A, 2B, and 3). An L1M4 truncated to a size of about 209 bp inserted within the L1HS301 locus over 60 MYA, as confirmed by the element's presence in prosimians. This L1M4 element belongs to an L1 subfamily that has an

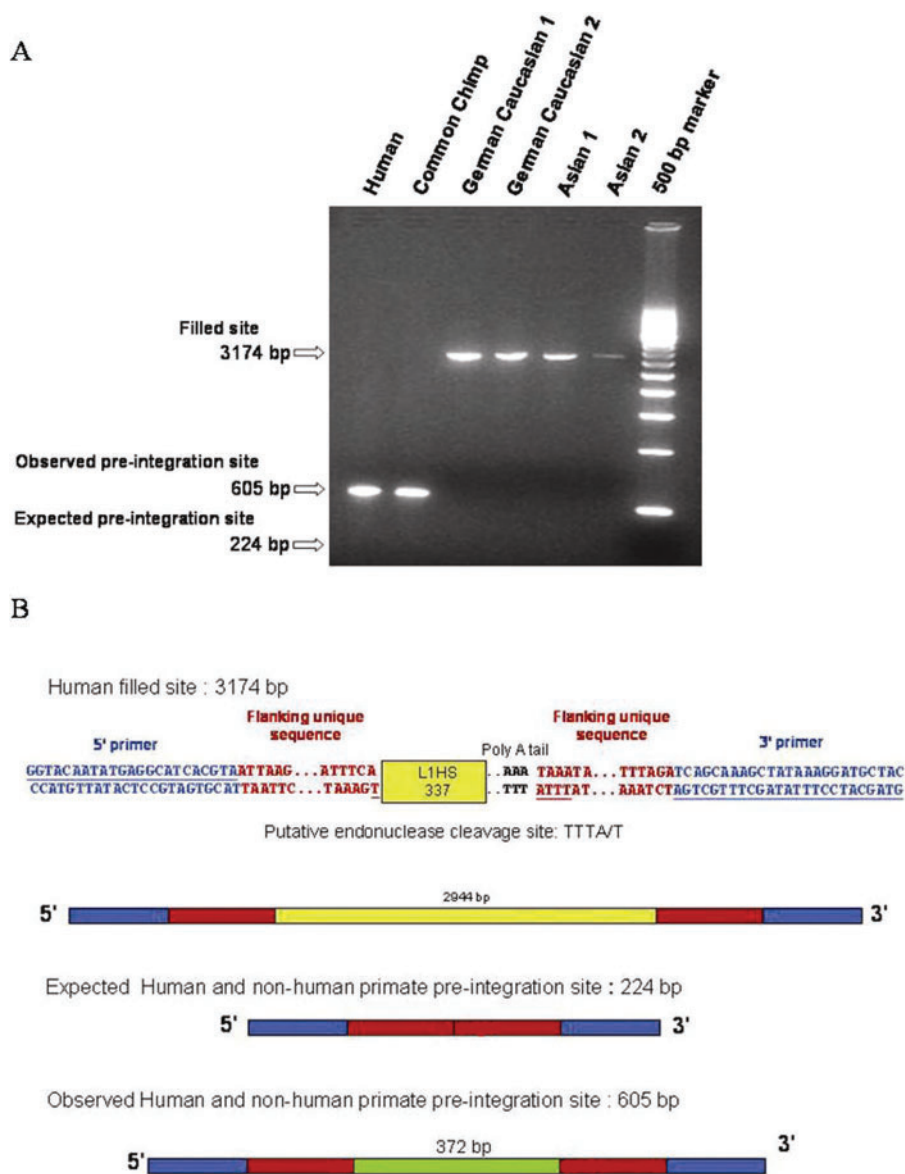


FIG. 1.—Analysis of L1HS337 locus. (A) An agarose gel chromatograph for the L1HS337 locus in five humans and a common chimpanzee. The calculated size of the preintegration site according to the GenBank sequence data is 224 bp. However, the actual preintegration site observed in some human individuals (heterozygotes and those homozygous $-/-$ for the L1 Ta insertion) and in nonhuman primates spanned a length of approximately 605 bp, as seen in the first two lanes above. Human individuals who are heterozygotes or who are homozygous $+/+$ for the presence of the L1 Ta insertion produce a 3,174-bp fragment identical to the GenBank sequence for the locus. (B) Deletion event at the L1HS337 locus. The first two schematic figures show the sequence of the L1HS 337 locus in human individuals possessing the L1 Ta element target site as indicated in red, and the L1HS 337 insertion followed by its 3' poly (A) tail as denoted in yellow. The expected human empty or preintegration site would be approximately 224 bp, with an endonuclease cleavage site of 5' TTTA/T. The observed preintegration site in both humans and nonhuman primates far exceeds the estimated PCR product, with a size of 605 bp. The sequence composition includes an additional 372 bp of nonrepetitive sequence (indicated in green), which is deleted as a result of the L1 integration.

average divergence from the L1 Ta subfamily of 22% (Smit et al. 1995), further indication that this element inserted into mammalian lineages before the primate radiation. After the divergence of prosimians approximately 60 MYA but before the divergence of the New World monkeys approximately 35 MYA, an Alu J integrated within the preexisting L1M4 at the locus. An additional mutation, a deletion approximately 100 bp in length occurring within the L1M4 element 3' of the Alu J integration site, is present

at the locus in the orangutan genome. Since PCR analysis of the green monkey locus did not produce an amplicon, we conclude that the deletion event occurred subsequent to the divergence of the New World monkeys but before the divergence of the Asian apes, placing the event between 35 and 15 MYA. Finally, a 1,085-bp truncated L1 Ta element inserted into a target site within the ancestral L1M4 element in the human lineage. Each successive insertion and deletion at the L1HS301 locus occurred within the

Table 1
Orthologous Loci Possessing Localized Sequence Variation Less Than 100 bp

L1Hs Ta Locus	Human	Chimpanzee	Pygmy Chimpanzee	Gorilla	Orangutan	Green Monkey	Owl Monkey	Galago
L1HS2	—	—	—	—	—	—	+21 bp	0
L1HS24	—	—	—	—	—	—	-10 bp	0
L1HS62	-62 bp	—	—	—	—	—	—	0
L1HS70	—	—	—	—	—	—	+24 bp	0
L1HS151	L1 Ta +/+	—	—	—	—	—	-13 bp (2 regions)	0
L1HS195	L1 Ta +/+	—	—	—	—	+28 bp	+28 bp	0
L1HS214	—	—	—	—	-38bp	0	0	0
L1HS314	—	—	—	—	—	0	+16 bp	0
L1HS315	L1 Ta +/+	—	—	—	—	+22 bp	0	0
L1HS443	L1 Ta +/+	—	—	+97 bp duplication	—	—	—	—
L1HS456	L1 Ta +/+	—	—	—	—	+18 bp	0	0
L1HS509	L1 Ta +/+	—	—	—	—	-13 bp	-13 bp	0

NOTE.—The dash (—) is used when corresponding PCR results agree with the computationally calculated L1 Ta preintegration locus as derived from human GenBank sequence; 0 denotes no PCR product observed at the locus; L1 Ta +/+ indicates a locus for which all human individuals are homozygous for the presence of the L1 Ta insertion; + indicates the addition of mobile element sequence or other sequence regions relative to the expected preintegration sequence; - indicates the absence of mobile element sequence or other sequence regions relative to the expected preintegration sequence or other primate sequences.

preexisting L1M4 element and was inherited in an IBD manner throughout primate evolution.

Parallel Independent Mobile Element Insertions

Six of 249 L1 Ta preintegration site loci analyzed (2.4%) have experienced the parallel insertion of one or more additional mobile elements at the same target site as the human L1 Ta insertion (L1HS45) and at sites ranging from 15 bp to 85 bp away in the other five instances (L1HS363, L1HS413, L1HS480, L1HS521, and L1HS561). Several examples of parallel independent Alu insertions within the same preintegration site sequence have been reported previously, in which an evolutionarily young Alu element inserts in the human genome into the same genomic region that contains an older Alu subfamily member in the owl monkey genome (Roy-Engel et al. 2002). Alu elements are the most abundant retroelements found in primate genomes, and they share the L1 retrotransposition machinery and therefore the same initial target site preference as L1 elements. Therefore, it is not surprising that five of these six parallel mobile element insertions involve the integration of Alu sequences. The parallel insertion events involving Alu elements that are reported here show an insertion bias towards the genomes of primate lineages that diverged at an evolutionary period when the rate of Alu amplification was as much as two orders of magnitude greater than it is at the present time (Roy-Engel et al. 2002). There was, however, one instance (L1HS561 locus) in which a parallel Alu insertion occurred in the gorilla genome. There was also a single example (L1HS45 locus) of a non-Alu related 5S rRNA

coding sequence inserting in parallel at an L1 Ta preintegration site locus in the owl monkey genome. It is interesting to note that the 5S rRNA DNA coding sequence found at the owl monkey locus retains an intact internal control region (ICR) that serves as the recognition site for transcription by polymerase III (Engelke et al. 1980; Pelham and Brown 1980), which may suggest that this is a recent event in the owl monkey genome or that the gene may retain the ability to be transcribed.

The L1HS480 locus (GenBank accession numbers AY246523 to AY246528) has been particularly receptive to multiple independent mobile element insertions over the course of primate evolution (figs. 3, 4A and 4B). A PCR assay was performed on an extensive nonhuman primate panel incorporating 14 different species to more accurately delineate the mode and tempo of these insertion events. The results of the PCR analysis indicate the presence of an Alu S/Y mosaic element exclusively in the owl monkey genome about 40 bp from the human L1 Ta element integration. In addition, the Old World monkeys (green monkey, pig-tailed macaque, and rhesus macaque) share an independent Alu Y insertion that integrated 16 bp from the L1 Ta target site. These two independent Alu insertions as well as the L1 Ta insertion in the human lineage are distinct, and they are easily distinguished from one another by their subfamily-specific mutations and unique target site duplications. This region on chromosome 4q may be particularly susceptible to mobile element insertions as a result of some undefined unique chromatin structure. However, detailed sequence analysis of 25 kb of flanking sequence on either side of this locus did not yield any coding sequences or any other distinguishing sequence

Table 2
Orthologous Loci Possessing Microsatellite/Low Complexity Sequence Variation

L1Hs Ta Locus	Human	Chimpanzee	Pygmy Chimpanzee	Gorilla	Orangutan	Green Monkey	Owl Monkey	Galago
L1HS284	176 bp	249 bp	239 bp	179 bp	247 bp	134 bp	139 bp	0
L1HS348	523 bp	497 bp	~523	523 bp	~523	467 bp	0	0
L1HS411	134 bp	114 bp	126 bp	124 bp	122 bp	129 bp	126 bp	127 bp

NOTE.—0 denotes no PCR product observed at the locus; ~ denotes approximate PCR product sizes based upon UV visualization of PCR products on agarose gel electrochromatographs in comparison to sequenced products at the same locus.

Table 3
Orthologous Loci Possessing Successive Mobile Element Insertions

L1Hs Ta Locus	Human	Chimpanzee	Pygmy Chimpanzee	Gorilla	Orangutan	Green Monkey	Owl Monkey	Galago
L1HS86	—	—	—	—	—	—AluY	—AluY	0
L1HS135	—	—	—	—	—	—AluY	—AluY	0
L1HS155	—	—	—	—	—L1PA3	—L1PA3	0	0
L1HS301	—	—	—	—	—	—	+100bp (L1M4)	+100bp;—AluJ
L1HS361	—	—	—	—	—	+MER65	0	0

NOTE.—The dash (—) is used when corresponding PCR results agree with the computationally calculated L1 Ta preintegration locus as derived from human GenBank sequence; 0 denotes no PCR product observed at the locus; + indicates the addition of mobile element sequence or other sequence regions relative to the expected preintegration sequence; — indicates the absence of mobile element sequence or other sequence regions relative to the expected preintegration sequence or other primate sequences.

features that would suggest that this locus provides a hotspot for mobile element insertion. The L1HS480 locus illustrates an extreme example of the potential pitfalls associated with assessing phylogenetic relationships based on a mobile element locus, since the two independent Alu insertions would not be distinguishable based upon size. The Alu-based insertion homoplasy of this locus is only evident after detailed sequence analysis.

Considering all six loci at which parallel independent insertions have occurred (L1HS45, L1HS363, L1HS413, L1HS480, L1HS521, and L1HS561), we observe seven potential insertion events out of 1,336 orthologous loci examined, an overall parallel insertion frequency of 0.52% (table 4). One event took place in the gorilla genome, another in the galago genome, three in the owl monkey genome, and two in the green monkey genome. The total number of L1 Ta loci that were successfully screened by PCR for each primate species is shown in table 4. For instance, 246 loci produced PCR amplicons in the gorilla, and only one of these represented a parallel independent insertion of another mobile element. Therefore, 1/246 or 0.41% of L1 Ta integration sites contain an independent retroelement insertion at their corresponding orthologous locus in the gorilla genome. Similarly, the percentage of L1 Ta loci containing parallel insertions in the green monkey genome is 2/191 or 1.05%, whereas the percentages in the owl monkey and galago genomes are 2.01% and 5%, respectively. Alternately, we may choose to discard the potentially paralogous L1HS521 locus as well as the L1HS45 locus containing the 5S rRNA coding sequence from further analysis due to their novelty. If we examine the frequencies of the remaining five parallel Alu insertions into L1 Ta orthologs in the various primate genomes, the frequency of L1 Ta integration sites with Alu elements in the green monkey genome is 0.53%, and in the owl monkey genome the frequency is 1.35%. Regardless of which calculations are considered, it is evident that the majority of independent parallel insertions in nonhuman primate L1 orthologs involve Alu elements, and the frequency of occurrence of these parallel Alu insertions parallels the rate of Alu amplification throughout primate evolution (Shen, Batzer, and Deininger 1991; Deininger and Batzer 1999). The only anomaly in our observations involves the relatively high level of independent mobile element insertions in galago (5%), but this is likely a statistical fluctuation attributable to the small number of L1 Ta preintegration loci that amplified in prosimian genomes.

Discussion

The retroelements that have had the most dramatic impact in shaping primate genomes are the L1 family of LINE elements and the Alu elements, their partner SINES (Prak and Kazazian 2000; Batzer and Deininger 2002). The copy and paste mechanism employed by these elements in their own amplification leads to an expansion in the size and composition of the genome, providing genomic markers that are thought to be homoplasy free characters. Studies of primate and mammalian orthologous sequences (Miyamoto, Slightom, and Goodman 1987; Ryan, Zielinski, and Dugaiczky 1991), as well as the subfamily hierarchy of accumulated mutations, confirm that the absence of retroelements is the ancestral state. In addition, there is no known mechanism for the precise deletion of these mobile elements, a phenomenon supported by the fossil sequence relics of retroelements belonging to subfamilies whose amplification occurred hundreds of millions of years ago (Goncalves, Duret, and Mouchiroud 2000; Deininger and Batzer 2002). These characteristics of LINES and SINES have rendered them extremely useful in reconstructing mammalian phylogenetic relationships, especially when they are used to deduce evolutionary relationships that are not suggested by morphological data (Ryan and Dugaiczky 1989; Verneau, Catzeflis, and Furano 1997; Hamdi et al. 1999; Nikaido et al. 2001). L1 and Alu elements have been equally powerful tools for the study of human evolution and human population dynamics (Jorde et al. 2000; Shedlock and Okada 2000; Carroll et al. 2001; Watkins et al. 2001; Myers et al. 2002; Bamshad et al. 2003).

Despite convincing examples of these retroelements' utility in deciphering evolutionary relationships with a complete lack of homoplasy, incidences of multiple independent insertions being found at identical or nearly identical sites in different species do exist. For example, two independent *mys* LTR-containing retrotransposon insertions have been identified at the same locus in rodent species (Cantrell et al. 2001). This phenomenon of parallel independent insertions has also resulted in the existence of an element belonging to the B1 family of SINES at the same insertion site in two different rodent species (Kass, Raynor, and Williams 2000). Three additional parallel independent Alu insertions have occurred at orthologous loci in the human and owl monkey genomes (Roy-Engel et al. 2002). Thus the homoplasy free nature of mobile elements has come into question (Hillis 1999).

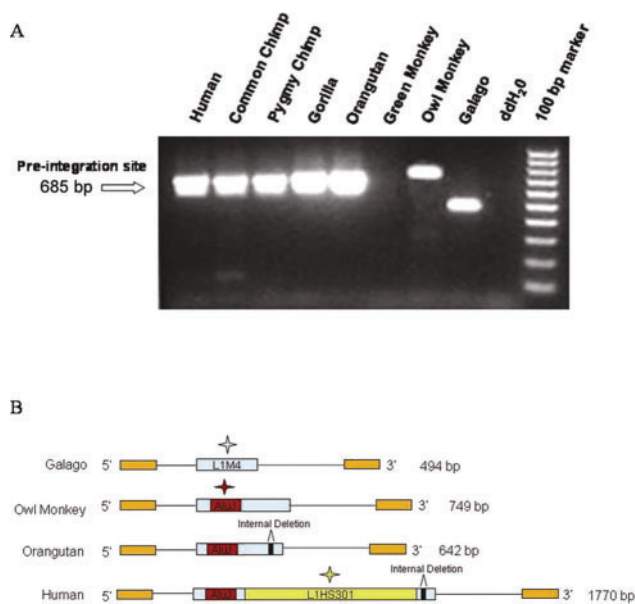


FIG. 2.—Variation within the L1HS301 preintegration site. (A) An agarose gel chromatograph of the PCR products derived from a phylogenetic analysis of the L1HS301 locus is shown. The observed human preintegration site contains the 685-bp product predicted computationally, and similar products result from PCR amplification in the common and pygmy chimpanzees, gorilla, and orangutan. The green monkey ortholog did not produce a PCR product. The owl monkey genome contains a larger product (749 bp) at the L1HS301 locus, and the galago genome contains a smaller product (494 bp). (B) The evolutionary history of the L1HS301 locus. A truncated 290-bp L1 element belonging to the L1M4 subfamily (represented in blue) inserted into the L1HS301 locus before the primate radiation. After the divergence of prosimians, an Alu J (represented in red) inserted into the primate lineage within the preexisting L1M4, an evolutionary event shown schematically in the 749-bp owl monkey locus. After the divergence of New World monkeys and before the divergence of the Asian apes (about 35 and 15 MYA), an internal deletion spanning about 100 bp (shown in black) occurred in the L1M4 element within the locus, as represented in the figure by the 642-bp orangutan sequence. Finally, a Ta L1 element L1HS301 (shown in yellow), inserted into the human genome within the ancestral L1M4 element between the Alu J subfamily member and the internal deletion.

It may be postulated that these examples of parallel independent retroelement insertions at orthologous loci combined with the instances reported here indicate that hotspots for mobile element insertion may exist at particular loci in related species. In order to explore this possibility, it was necessary to calculate whether the number of parallel insertion events found at L1 Ta integration sites differed significantly from the number of parallel insertion events that would be expected to occur by chance. Considering an average target site sequence length of 175 (225 minus an average of 50 bp consisting of the flanking primer binding sequences), a total of 233,800 bp of potential target site DNA was successfully screened in seven nonhuman primate species, including the common chimpanzee, pygmy chimpanzee, gorilla, orangutan, green monkey, owl monkey, and galago. Assuming a total genome size of approximately 3×10^9 bp and an average target site of 175 bp, we would expect about 17 million target sites per primate genome if we also assume that mobile element integration is entirely random. Combining the estimated L1 copy number of about

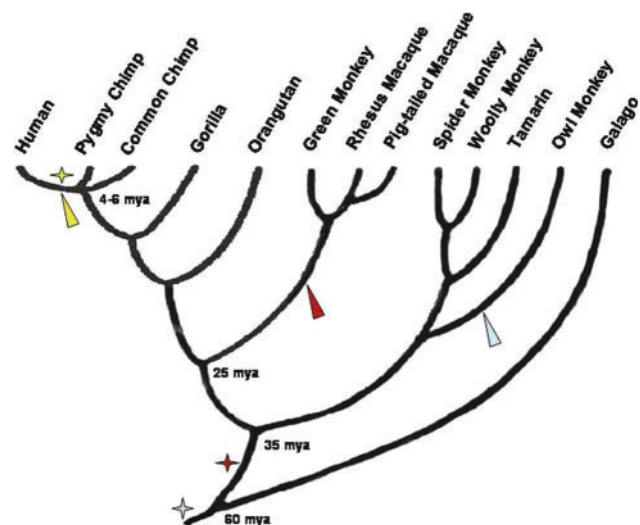
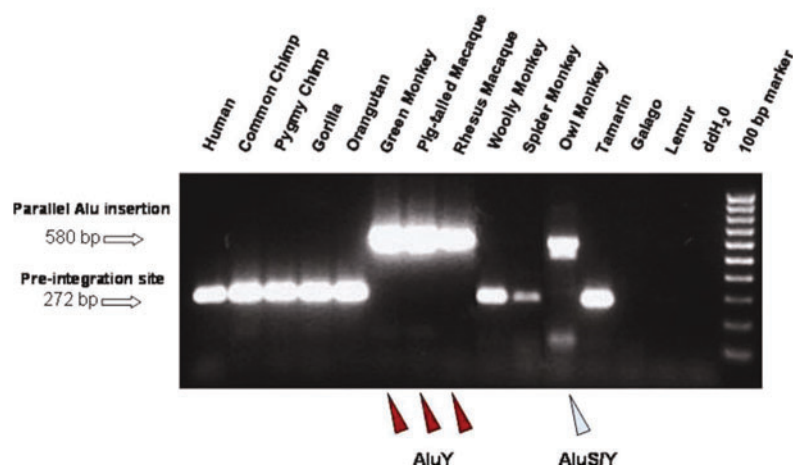


FIG. 3.—Schematic representation of the primate evolutionary tree. The estimated time of divergence for selected branches is labeled in millions of years before the present. The stars indicate successive mobile element insertion events at the L1HS301 locus. The blue star represents an L1M4 integration occurring before the primate radiation, the red star indicates an Alu J insertion occurring after the prosimian divergence but before the divergence of New World monkeys, and the yellow star represents the L1 Ta element insertion (L1HS301) in the human genome within the last 4 to 6 Myr. The elongated triangles represent independent parallel mobile element insertion events occurring at the L1HS480 locus in various primate species. The blue triangle represents the Alu S/Y mosaic, which is found exclusively in the owl monkey genome; the red triangle indicates an AluY insertion occurring in the Old World monkey lineage such that the insertion appears in the green monkey, rhesus macaque, and pig-tailed macaque genomes; and the yellow triangle indicates the L1 Ta element insertion (L1HS480) in the human genome.

516,000 and estimated Alu copy number of 1.09 million in the human genome, we arrive at 1.606 million total mobile element insertions distributed among 17 million target sites (International Human Genome Sequencing Consortium 2001). The probability of finding either an L1 or Alu element within any randomly chosen 175 bp target site within the human genome would be about 9.45% (1.606/17). Using this figure, we can predict that two independent insertions would share the same target site at orthologous loci of primate genomes of comparable size and repeat copy number approximately 0.89% (0.0945^2) of the time due to chance alone. This figure is in good agreement with the average percentage of parallel Alu insertion events at L1 Ta element integration loci over seven nonhuman primate species of 1.04%, a figure calculated after discarding the potentially paralogous green monkey locus (L1HS521). Therefore, it appears that the observed frequency of independent parallel mobile element insertions at the orthologous loci of L1 Ta elements is not substantially greater than what would be expected to occur by chance alone.

The observation of three independent insertion events occurring at the L1HS480 locus calls into question the randomness of independent parallel insertion events, since the likelihood of this happening by chance alone is extremely low ($\sim 0.084\%$). Perhaps this locus falls within an area prone to genomic instability related to retrotransposition events due to chromatin configuration,

A



B

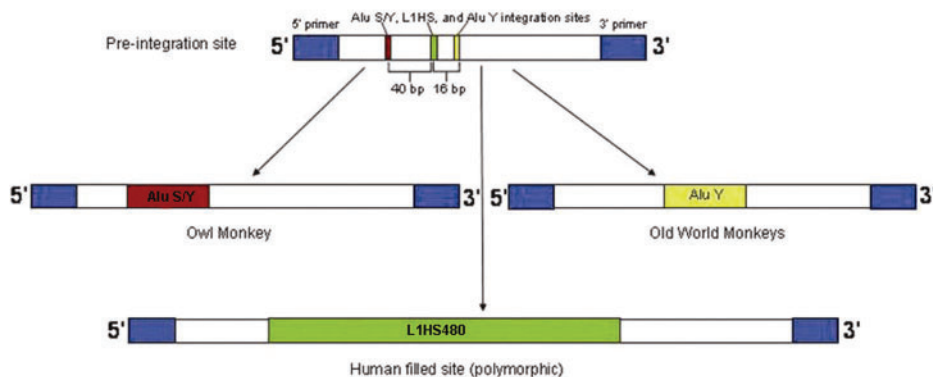


FIG. 4.—Parallel insertions at the L1HS480 locus. (A) Diagram depicting multiple insertions within the last 35 Myr of primate evolution. The figure shows an agarose gel chromatograph of the PCR products resulting from amplification at the L1HS480 locus in 14 distinct primate species. The estimated 272-bp preintegration product is found in the human, common chimpanzee, pygmy chimpanzee, gorilla, orangutan, woolly monkey, spider monkey, and tamarin genomes. An independent Alu Y integration (indicated by the red triangles) in the lineage that led to the Old World monkeys results in a 588-bp amplicon in the orthologous locus of the green monkey, pig-tailed macaque, and rhesus macaque. An additional independent insertion of an Alu S/Y mosaic (indicated by the blue triangle) occurred exclusively in the owl monkey genome after divergence from other New World monkeys, creating a 588-bp product in the owl monkey genome. (B) Multiple parallel insertions in the L1HS480 locus throughout primate evolution. The locus labeled “preintegration site” at the top of the figure represents the unique sequence constituting an empty or “unfilled” site at the L1HS480 locus in the galago, orangutan, gorilla, pygmy chimpanzee, and common chimpanzee genomes. This preintegration site is also found in human individuals who are homozygous for the absence of the L1HS480 insertion as well as in human individuals who are heterozygotes for L1 insertion presence at this locus. The narrow colored bars within the figure indicate the relative orientations and distances between the insertion sites of the various mobile elements. The integration site of the Alu S/Y mosaic is indicated in red. The L1HS480 integration site occurs 40 bp 3' of the Alu S/Y element, and it is indicated in green. The Alu Y integration site occurs an additional 16 bp 3' of the L1HS insertion site, and it is represented in yellow. The three independent insertions each have their own respective target site duplications characteristic of three separate retrotransposition events.

allowing the retrotransposition machinery comparatively easy access to this sequence region. However, the overall level of mobile element insertion site homoplasy will vary both as a result of the relative rates of retrotransposition in different genomes and the length of time since the divergence of species. Thus the longer the evolutionary time frame involved, the greater the opportunity for insertion site homoplasy.

It is also important to note that none of the cases of insertion homoplasy involved multiple L1 insertions that occurred in parallel in different primate genomes. The reason for this difference in insertion site homoplasy between Alu and L1 elements is unclear. However, it

may be related to genome structural constraints imposed on L1 insertion events as a result of size differences between Alu and L1 elements. Alternatively, it may be the result of a reduced rate of amplification of L1 elements in some of the nonhuman primate genomes in which Alu elements appear to be currently undergoing rapid retrotransposition (e.g., owl monkey). In addition, the likelihood for precise insertion site homoplasy with L1 elements is lower since this type of event would involve two independent L1 insertions at the same genomic location along with two variable truncations of newly integrated L1 elements to about the same length (a reasonably rare event).

Table 4
Independent Parallel Insertions of Mobile Elements at L1 Ta Preintegration Loci

L1Hs Ta locus	Human	Chimpanzee	Pygmy Chimpanzee	Gorilla	Orangutan	Green Monkey	Owl Monkey	Galago
L1HS45	—	—	—	—	—	—	+5S rRNA	0
L1HS363	—	—	—	—	—	—	—	+ type II Alu
L1HS413	—	—	—	—	—	—	+AluSc	0
L1HS480	—	—	—	—	—	+AluY	+AluS/Y	0
L1HS521	—	—	—	—	—	+AluY;—	—	—
L1HS561	—	—	—	+AluY	—	—	—	0
Total Loci Analyzed	249	248	248	246	234	191	149	20

NOTE.—The dash (—) is used when corresponding PCR results agree with the computationally calculated L1 Ta preintegration locus as derived from human GenBank sequence; 0 denotes no PCR product observed at the locus; + indicates the addition of mobile element sequence or other sequence regions relative to the expected preintegration sequence.

The other identified sources of variability in the L1 Ta element preintegration sites in nonhuman primates can be attributed to common mutation processes that generate sequence diversity. A single mutational event during primate evolution, resulting from the process of replication, recombination, gene conversion, or repair, is sufficient to account for the sequence variation observed at 12 L1 Ta loci. Similarly, three loci differ among nonhuman primates due to variable microsatellite sequence lengths caused by slippage during the replication process. These minor variants spanning less than 100 bp account for 50% (15/30) of the sequence variation observed at the L1 Ta preintegration site loci.

The three definitive examples of deletions caused by the L1 insertions represent 8.11% (3/37) of the 30 insertion sites sequenced in this study in combination with seven additional insertion sites sequenced previously (Morrish et al. 2002). If we include the two putative single-base deletion events, the estimated percentage of L1 insertions that cause target site deletions during retrotransposition jumps to 13.5%. Although lower than expected, these figures are in relatively good agreement with previous estimates of 8/37 (21.62%) of retrotransposition events in cultured human cells causing deletions at the target site, considering that some deletion events may become obscured or eliminated at the sequence level over time via negative selection, mutation, recombination, or repair (Gilbert, Lutz-Prigge, and Moran 2002).

Five of the 30 variable preintegration loci analyzed (16.67%) consisted of successive mobile element insertion events that were inherited throughout the primate lineage as an integral part of the preintegration site into which the L1 Ta element subsequently inserted. The sequences of these orthologous loci allow us to trace the succession of mobile element integrations at a single locus throughout primate evolution. These clustered insertions appear to support the notion that regions of previously existing mobile elements such as Alu poly (A) tails and L1 A/T-rich sequence, can serve as potential target sites for subsequent mobile element integration. The L1HS301 locus provides an interesting examination of the alteration of a genomic locus that can occur from multiple mobile element insertions and sequence deletions within the interior of preexisting mobile elements. Sequence analysis may be necessary to confirm the sequence composition of loci containing adjacent mobile elements, particularly when PCR analysis indicates variability in the

sizes of preintegration loci across several distantly related species.

In conclusion, about 12% of L1 Ta element insertion sites show sequence size and content variability at their orthologous loci in nonhuman primate genomes. About half of these variants are minor, involving short sequence tracts of less than 100 bp. Target site deletions occur as a result of approximately 10% of the L1 insertion events in the genome, and these deletions range from a single nucleotide to over 25 kb. Insertion site sequence architecture may also be altered in approximately 17% of variable L1 integration sites over the course of evolution due to the successive accumulation of multiple mobile elements within a small genomic region. Parallel independent insertions of mobile elements into orthologous loci in nonhuman primates are an additional source of sequence variation. The rates of parallel insertions vary predictably according to species and mobile element type due to differential amplification rates over the course of primate evolution. These events can potentially introduce homoplasmy into retroelement-based phylogenetic and population genetic data that is otherwise homoplasmy free. Fortunately, these events are easily discernible as variably truncated product sizes when L1 elements are involved in the parallel insertion events. Although parallel independent Alu insertions are more difficult to detect without sequence analysis, the overall percentage of parallel mobile element insertion events is relatively low, affecting only seven out of 1,336 or 0.52% of loci examined. These occurrences appear to be random and exceedingly rare, even when considering primate species that are separated by evolutionary time periods on the order of 50 to 60 Myr. In addition, no independent mobile element insertion events of any type have been found at orthologous loci in human and chimpanzee species, reaffirming the utility of L1 and Alu element integrations as essentially homoplasmy free characters well suited to the study of population genetics and phylogenetic relationships within closely related species.

Acknowledgments

This research was supported by Louisiana Board of Regents Millennium Trust Health Excellence Fund HEF (2000-05)-05, (2000-05)-01, and (2001-06)-02 (M.A.B.) and National Science Foundation grants BCS-0218338 (M.A.B.) and BCS-0218370 (L.B.J.).

Literature Cited

- Ausubel, F. M., R. Brent, M. E. Kingston, D. D. Moore, and J. G. Seidman. 1987. *Current Protocols in Molecular Biology*. Wiley, New York.
- Bamshad, M. J., S. Wooding, W. S. Watkins, C. T. Ostler, M. A. Batzer, and L. B. Jorde. 2003. Human population genetic structure and inference of group membership. *Am. J. Hum. Genet.* **72**:578–589.
- Batzer, M. A., and P. L. Deininger. 2002. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* **3**:370–379.
- Batzer, M. A., C. W. Schmid, and P. L. Deininger. 1993. Evolutionary analyses of repetitive DNA sequences. *Methods Enzymol.* **224**:213–232.
- Boeke, J. D., and O. K. Pickeral. 1999. Retroshuffling the genomic deck. *Nature* **398**:108–109, 111.
- Boissinot, S., P. Chevret, and A. V. Furano. 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol. Biol. Evol.* **17**:915–928.
- Cantrell, M. A., B. J. Filanoski, A. R. Ingermann, K. Olsson, N. DiLuglio, Z. Lister, and H. A. Wichman. 2001. An ancient retrovirus-like element contains hot spots for SINE insertion. *Genetics* **158**:769–777.
- Carrall, M. L., A. M. Roy-Engel, S. V. Nguyen et al. (16 co-authors). 2001. Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. *J. Mol. Biol.* **311**:17–40.
- Cost, G. J., and J. D. Boeke. 1998. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry* **37**:18081–18093.
- Deininger, P. 1993. Induction of DNA rearrangement and transposition. *Proc. Natl. Acad. Sci. USA* **90**:3780–3781.
- Deininger, P. L., and M. A. Batzer. 1999. Alu repeats and human disease. *Mol. Genet. Metab.* **67**:183–193.
- . 2002. Mammalian retroelements. *Genome Res.* **12**:1455–1465.
- Deininger, P. L., M. A. Batzer, C. A. Hutchison, 3rd, and M. H. Edgell. 1992. Master genes in mammalian repetitive DNA amplification. *Trends Genet.* **8**:307–311.
- Engelke, D. R., S. Y. Ng, B. S. Shastry, and R. G. Roeder. 1980. Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell* **19**:717–728.
- Fanning, T. G., and M. F. Singer. 1987. LINE-1: a mammalian transposable element. *Biochim. Biophys. Acta* **910**:203–212.
- Feng, Q., J. V. Moran, H. H. Kazazian, Jr., and J. D. Boeke. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**:905–916.
- Gilbert, N., S. Lutz-Prigge, and J. Moran. 2002. Genomic deletions created upon LINE-1 retrotransposition. *Cell* **110**:315–325.
- Goncalves, I., L. Duret, and D. Mouchiroud. 2000. Nature and structure of human genes that generate retropseudogenes. *Genome Res.* **10**:672–678.
- Goodier, J. L., E. M. Ostertag, and H. H. Kazazian, Jr. 2000. Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* **9**:653–657.
- Goodman, M. 1999. The genomic record of humankind's evolutionary roots. *Am. J. Hum. Genet.* **64**:31–39.
- Hamdi, H., H. Nishio, R. Zielinski, and A. Dugaiczky. 1999. Origin and phylogenetic distribution of Alu DNA repeats: irreversible events in the evolution of primates. *J. Mol. Biol.* **289**:861–871.
- Hillis, D. M. 1999. SINEs of the perfect character. *Proc. Natl. Acad. Sci. USA* **96**:9979–9981.
- Hohjoh, H., and M. F. Singer. 1996. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.* **15**:630–639.
- Jorde, L. B., W. S. Watkins, M. J. Bamshad, M. E. Dixon, C. E. Ricker, M. T. Seielstad, and M. A. Batzer. 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66**:979–988.
- Jurka, J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci. USA* **94**:1872–1877.
- Kass, D. H., M. E. Raynor, and T. M. Williams. 2000. Evolutionary history of B1 retroposons in the genus *Mus*. *J. Mol. Evol.* **51**:256–264.
- Kazazian, H. H., Jr. 2000. Genetics: L1 retrotransposons shape the mammalian genome. *Science* **289**:1152–1153.
- Kazazian, H. H., Jr., and R. G. Cotton. 2001. In memoriam. *Hum. Mutat.* **18**:461.
- Kazazian, H. H., Jr., and J. V. Moran. 1998. The impact of L1 retrotransposons on the human genome. *Nat. Genet.* **19**:19–24.
- Kazazian, H. H., Jr., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips, and S. E. Antonarakis. 1988. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332**:164–166.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921.
- Levinson, G., and G. A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**:203–221.
- Luan, D. D., M. H. Korman, J. L. Jakubczak, and T. H. Eickbush. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**:595–605.
- Malik, H. S., and T. H. Eickbush. 1999. Retrotransposable elements R1 and R2 in the rDNA units of *Drosophila mercatorum*: abnormal abdomen revisited. *Genetics* **151**:653–665.
- Mathews, L. M., S. Y. Chi, N. Greenberg, I. Ovchinnikov, and G. D. Swergold. 2003. Large differences between LINE-1 amplification rates in the human and chimpanzee lineages. *Am. J. Hum. Genet.* **72**:739–748.
- Meischl, C., M. Boer, A. Ahlin, and D. Roos. 2000. A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. *Eur. J. Hum. Genet.* **8**:697–703.
- Miyamoto, M. M., J. L. Slightom, and M. Goodman. 1987. Phylogenetic relations of humans and African apes from DNA sequences in the psi eta-globin region. *Science* **238**:369–373.
- Moran, J. V., R. J. DeBerardinis, and H. H. Kazazian, Jr. 1999. Exon shuffling by L1 retrotransposition. *Science* **283**:1530–1534.
- Morrish, T. A., N. Gilbert, J. S. Myers, B. J. Vincent, T. D. Stamato, G. E. Taccioli, M. A. Batzer, and J. V. Moran. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* **31**:159–165.
- Myers, J. S., B. J. Vincent, H. Udall et al. (12 co-authors). 2002. A comprehensive analysis of recently integrated human Ta L1 elements. *Am. J. Hum. Genet.* **71**:312–326.
- Nikaido, M., F. Matsuno, H. Hamilton et al. (11 co-authors). 2001. Retroposon analysis of major Cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. *Proc. Natl. Acad. Sci. USA* **98**:7384–7389.
- Ovchinnikov, I., A. Rubin, and G. D. Swergold. 2002. Tracing the LINES of human evolution. *Proc. Natl. Acad. Sci. USA* **99**:10522–10527.

- Pelham, H. R., and D. D. Brown. 1980. A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. *Proc. Natl. Acad. Sci. USA* **77**:4170–4174.
- Prak, E. T., and H. H. Kazazian, Jr. 2000. Mobile elements and the human genome. *Nat. Rev. Genet.* **1**:134–144.
- Rothbarth, K., A. Hunziker, H. Stammer, and D. Werner. 2001. Promoter of the gene encoding the 16 kDa DNA-binding and apoptosis-inducing C1D protein. *Biochim. Biophys. Acta* **1518**:271–275.
- Roy-Engel, A. M., M. L. Carroll, M. El-Sawy, A. H. Salem, R. K. Garber, S. V. Nguyen, P. L. Deininger, and M. A. Batzer. 2002. Non-traditional Alu evolution and primate genomic diversity. *J. Mol. Biol.* **316**:1033–1040.
- Ryan, S. C., and A. Dugaiczky. 1989. Newly arisen DNA repeats in primate phylogeny. *Proc. Natl. Acad. Sci. USA* **86**:9360–9364.
- Ryan, S. C., R. Zielinski, and A. Dugaiczky. 1991. Structure of the gorilla alpha-fetoprotein gene and the divergence of primates. *Genomics* **9**:60–72.
- Salem, A. H., J. S. Myers, A. C. Otieno, W. S. Watkins, L. B. Jorde, and M. A. Batzer. 2003. LINE-1 pre-Ta elements in the human genome. *J. Mol. Biol.* **326**:1127–1146.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
- Schlotterer, C., and D. Tautz. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20**:211–215.
- Schwahn, U., S. Lenzner, J. Dong et al. (16 co-authors). 1998. Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nat. Genet.* **19**:327–332.
- Shedlock, A. M., and N. Okada. 2000. SINE insertions: powerful tools for molecular systematics. *Bioessays* **22**:148–160.
- Sheen, F. M., S. T. Sherry, G. M. Risch, M. Robichaux, I. Nasidze, M. Stoneking, M. A. Batzer, and G. D. Swergold. 2000. Reading between the LINES: human genomic variation induced by LINE-1 retrotransposition. *Genome Res.* **10**:1496–1508.
- Shen, M. R., M. A. Batzer, and P. L. Deininger. 1991. Evolution of the master Alu gene(s). *J. Mol. Evol.* **33**:311–320.
- Singer, M. F. 1982. SINES and LINES: highly repeated short and long interspersed sequences in mammalian genomes. *Cell* **28**:433–434.
- Singer, M. F., V. Krek, J. P. McMillan, G. D. Swergold, and R. E. Thayer. 1993. LINE-1: a human transposable element. *Gene* **135**:183–188.
- Skowronski, J., T. G. Fanning, and M. F. Singer. 1988. Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol. Cell Biol.* **8**:1385–1397.
- Skowronski, J., and M. F. Singer. 1986. The abundant LINE-1 family of repeated DNA sequences in mammals: genes and pseudogenes. *Cold Spring Harb. Symp. Quant. Biol.* **51**(pt 1): 457–464.
- Smit, A. F. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**:657–663.
- Smit, A. F., G. Toth, A. D. Riggs, and J. Jurka. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.* **246**:401–417.
- Verneau, O., F. Catzeflis, and A. V. Furano. 1997. Determination of the evolutionary relationships in *Rattus sensu lato* (Rodentia : Muridae) using L1 (LINE-1) amplification events. *J. Mol. Evol.* **45**:424–436.
- Watkins, W. S., C. E. Ricker, M. J. Bamshad, M. L. Carroll, S. V. Nguyen, M. A. Batzer, H. C. Harpending, A. R. Rogers, and L. B. Jorde. 2001. Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *Am. J. Hum. Genet.* **68**:738–752.
- Wei, W., N. Gilbert, S. L. Ooi, J. F. Lawler, E. M. Ostertag, H. H. Kazazian, J. D. Boeke, and J. V. Moran. 2001. Human L1 retrotransposition: *cis* preference versus trans complementation. *Mol. Cell Biol.* **21**:1429–1439.
- Woods-Samuels, P., C. Wong, S. L. Mathias, A. F. Scott, H. H. Kazazian, Jr., and S. E. Antonarakis. 1989. Characterization of a nondeleterious L1 insertion in an intron of the human factor VIII gene and further evidence of open reading frames in functional L1 elements. *Genomics* **4**:290–296.
- Yang, Z., D. Boffelli, N. Boonmark, K. Schwartz, and R. Lawn. 1998. Apolipoprotein(a) gene enhancer resides within a LINE element. *J. Biol. Chem.* **273**:891–897.

David Goldstein, Associate Editor

Accepted April 11, 2003