Characterization of a third major SINE family of repetitive sequences in the galago genome

Gary R.Daniels* and Prescott L.Deininger[1,2]
Department of Biochemistry, College of Medicine, University of South Alabama, Mobile, AL 36688, [1]Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, 1901 Perdido Street, New Orleans, LA 70112 and [2]Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, New Orleans, LA 70121, USA

Received October 31, 1990; Revised and Accepted March 12, 1991 EMBL accession nos+

ABSTRACT

In addition to the Alu family of short interspersed repetitive DNA elements (SINEs), we have previously characterized one other repetitive DNA family (Type II) in the prosimian, *Galago crassicaudatus*. We present here a detailed analysis of seventeen members of a third galago SINE family designated as the Monomer family. Both the Monomer and Type II families are shown to be specific for the galago genome as compared to other primates, including another prosimian, the lemur. Moreover, *in vitro* transcription of galago SINEs suggests that the Monomer and Type II families have appreciably stronger RNA polymerase III promoters than does the Alu family. This agrees with the promoter sequence for each of these SINE families, in that the Monomer and Type II family promoters are more closely related to the RNA polymerase III promoter consensus sequence than is the Alu family promoter. These promoter strength analyses also correlate with copy number and sequence divergence analyses, which suggests that the SINE families with the strongest promoters have been amplified most recently in the galago genome.

INTRODUCTION

Short interspersed repetitive DNA elements (SINEs) are a common feature of all mammalian genomes (1,2,3). Since there are different types of mammalian SINEs, these repetitive DNA elements have been grouped into families of related sequences (4). The most extensively characterized SINE family is known as the human Alu family, which exists as 500,000 repetitive DNA elements interspersed throughout the haploid human genome (5). The human Alu family of DNA elements corresponds to a 300 base pair (bp) dimeric structure composed of two related and tandemly arranged 133 bp monomeric units, with an additional 31 bps located in the middle of the right-half monomeric sequence (6). This dimeric structure of the Alu family has also been found in the genomes of all primates studied to date, including the prosimian, galago (7). The galago Alu family (also known as the Type I family) is closely related to the human Alu family and shows considerable sequence identity to the human dimer throughout its structure. In addition to the Alu family, there are two other major SINE families that have been discovered in the galago genome (8). One of these galago SINE families (the Monomer family) is not related to the Alu family, but does show a striking relatedness in sequence to methionine transfer RNA (tRNA) genes, suggesting that the Monomer family was derived from this tRNA gene (9). The other major galago SINE family (the Type II family) is a chimeric element composed of sequences closely related to the Monomer family in the left-half of its structure and sequences almost identical to the Alu (or Type I) family in the right-half of its structure. Thus, it appears that there was an ancestral fusion of a Monomer family element with an Alu family member that created an independent Type II DNA element in the galago genome (10). Figure 1 depicts a proposed mechanism for the formation of the Type II family in galago from pre-existing Monomer and Alu family elements. The schematic diagram also demonstrates the structural relationship between the three major galago SINE families (Figure 1).

All of the galago SINE families are thought to have been amplifed and dispersed throughout the genome by a transposition mechanism often referred to as retroposition (11). This mechanism involves the transposition of DNA elements through an RNA intermediate which is transcribed from the internal RNA polymerase III promoter of the SINE (12,13). As shown in Figure 1, retroposition also requires reverse transcription to generate SINE cDNAs that are subsequently integrated into the genome producing short direct repeats that flank the SINE. Although repetitive elements other than SINEs, such as LINEs and processed pseudogenes, can also undergo the retroposition process (2,3), it is clear that SINEs are particularly efficient at increasing their copy number in the genome by this process. Several factors may be important for the efficient amplification of SINEs, including the presence of an internal RNA polymerase

* To whom correspondence should be addressed

+ X55906, X00097, X00116, X00108, X00100, X03322, X00102, X00111, X03327, X03328, X55920, X03329, X55910, X03331-38, X55915

III promoter to direct the formation of SINE transcripts and the presence of an internal oligo(dA)-rich sequence near the 3' terminus to allow for productive self-priming of SINE transcripts during reverse transcription (12,13). Moreover, since most SINE families are derived from parental genes that are not highly amplified (for example, the $tRNA^{Met}$ gene for the Monomer family), mutational and/or structural modifications of the parental gene appear to be needed to increase the efficiency of the retroposition process before the SINE can be amplified (2,9). Finally, the physical location of a SINE in the genome may affect the amplification rate as flanking sequences are known to be involved in the regulation of some RNA polymerase III promoters *in vivo* (3,4).

The evolutionary relationships of SINE families also provides some information about the rate at which SINEs are amplified in the genome (2). In the case of the galago SINE families, it is clear that the Alu family is the most ancient of the three families. The Monomer family must also have pre-dated the development of the chimeric Type II family, since it was needed for the formation of the first Type II element. In order to more clearly evaluate the evolutionary relationship of the three major galago SINE families, we have characterized numerous additional Monomer family SINEs and compared their transcriptional properties to that of the Type II and Alu (Type I) families. Here, we describe some of the factors that may have influenced the amplification and evolution of these related galago SINE families.

MATERIALS AND METHODS

Isolation of Monomer family clones

Several Monomer family clones (GAL 2,32,38,39) characterized in this study have been isolated from a M13mp8 library of *Galago crassicaudatus* DNA as has been described previously (8). Additional Monomer family clones (GSE 7,9,12,18,20,32,36,37,40,41,43, 55) were isolated from a M13mp8 library of *Galago senegalensis* DNA using the methods described below. Genomic *Galago senegalensis* DNA was isolated from cultured primary fibroblast cells (supplied by Dr. P. Welsh, Duke University) by standard procedures (14). The genomic DNA was cleaved with the restriction endonuclease *Rsa*I, size-fractionated by agarose gel electrophoresis to obtain fragments of about 500 bps in length, and cloned into M13mp8 by blunt-end ligation as previously described (7,8). The *Galago senegalensis* M13mp8 library was then plated (15) and screened by plaque hybridization (16), using the insert from the Monomer family clone GAL 39 as the hybridization probe. Positive clones were then isolated and phage DNA was prepared from one ml cultures for DNA sequencing (17). All DNA sequencing was accomplished by the dideoxynucleotide chain terminating procedure of Sanger et al. (18) as modified for incorporation of [α-^{35}S]dATP into the nascent chain (19).

In vitro transcriptional analysis

The template for the *in vitro* transcription reactions was supercoiled replicative form (RF) DNA prepared from M13 clones characterized previously (7,8) and in this study as described above. The RF DNA was prepared using the alkaline lysis procedure (20) on bacteria infected with a single recombinant M13mp8 phage (21). All RF DNAs were purified by two rounds of centrifugation to equilibrium in CsCl-ethidium bromide gradients (14). The soluble extracts used for the transcription reactions were prepared as S100 extracts from HeLa cells grown in spinner flasks as described by Dignam et al. (22). A series of titrations of the DNA template concentration were carried out to optimize the transcription reactions with the S100 extract. All transcription reactions and RNA extraction conditions were similar to those we have described previously using [α-^{32}P]UTP as a source of label for transcription products (23). The standard reaction contained 12 mM HEPES, pH 7.9, 12% (v/v) glycerol, 7 mM $MgCl_2$, 60 mM KCl, 0.2 mM ethylenediamine tetraacetate (EDTA), 0.3 mM dithiothreitol, 0.5 mM each of ATP, GTP, CTP, 25 μM [α-^{32}P]UTP (2–4 Ci/mmole), 0.5 μg/ml α-amanitin, 1 μg of DNA, and 7 μl of S100 extract in a final volume of 25 μl. Transcription reactions were initiated by the addition of the S100 extract and the reaction was incubated at 30°C for 60 min. After incubation, the transcription reactions were terminated by the addition of 125 μl of a solution containing 8 M urea, 1% SDS, 10 mM EDTA, and 25 μg yeast tRNA and the labelled RNA was extracted several times with an equal volume of phenol:chloroform (1:1) to remove protein (24). The RNA was then isolated by precipitation with 95% ethanol and centrifugation to collect the pellet. The labelled transcripts were then dissolved in 10 μl of 98% formamide and separated on 6% polyacrylamide gels containing 7 M urea. Electrophoresis was continued until the bromophenol blue tracking dye reached the bottom of the gel. The position of specific transcripts in the gel was determined by autoradiography.

S1 nuclease mapping analysis

S1 nuclease mapping studies were carried out to map the 5' ends of the transcripts for some of the galago Monomer (GAL 2 and 39) and Type II (GAL 5 and 25) SINE family clones. For these experiments, transcripts were synthesized and purified as described in the preceding section except that [α-^{32}P]UTP was replaced with cold UTP in the transcription reaction. A 5'-end labelled DNA probe was prepared for each galago SINE subjected to this analysis by cleaving the recombinant M13mp8 clone at a restriction site within the repetitive DNA element, labelling the 5' ends with polynucleotide kinase (25) and [γ-^{32}P]ATP (3000 Ci/mmole), and then cleaving the M13mp8 vector at a site upstream of the cloned repetitive DNA insert. The appropriate 5'-end labelled SINE fragment was then isolated from a denaturing 7 M urea-6% polyacrylamide gel. The reaction conditions used for annealing the 5'-end labelled probe to the transcript and conducting the S1 nuclease mapping analysis were the same as described previously (26).

Determination of SINE family copy number

Because the Type II family has sequence elements related to those in both the Alu (Type I) family and the Monomer family, direct measurements of the copy numbers of these SINE families in galago genomic DNA by hybridization analysis is not possible. We have attempted to overcome this problem by screening 602 random M13mp8 clones from the *Galago senegalensis* library, with probes specific for each galago SINE family. The probes used were all single-stranded DNA produced by primed synthesis reactions. In the first step of this procedure, 2 μg of single-stranded DNA from the desired recombinant M13mp8 clone was annealed to an excess of the universal sequencing primer to provide a template for DNA synthesis (21). The primer was extended in a reaction mixture containing 10 mM Tris (pH 7.4), 6 mM $MgCl_2$, 50 μM each of dGTP, dCTP, dTTP and 20 μCi of [α-^{32}P]dATP and two units of the Klenow fragment of DNA

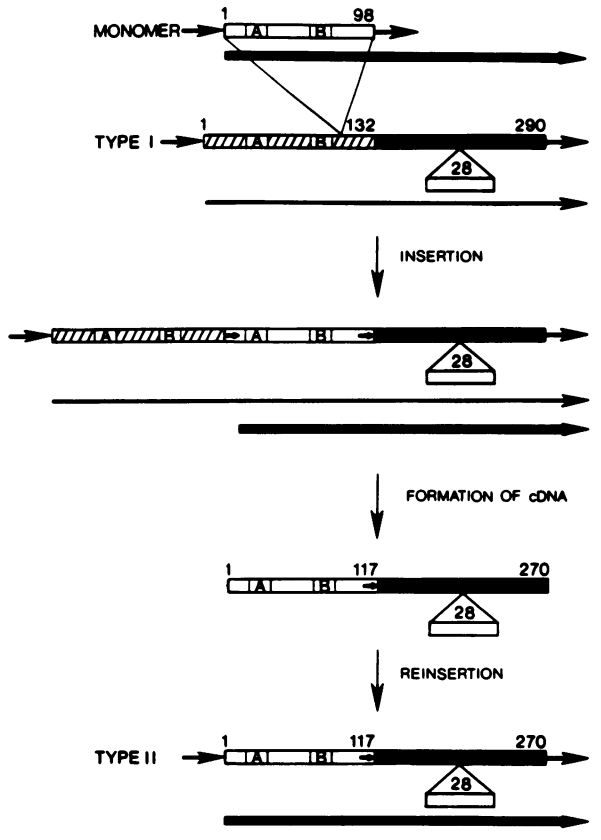

Figure 1. Proposed mechanism for the formation of the Type II SINE family. The schematic diagram shows a putative mechanism for the formation of the Type II family and the structural relatedness of the three galago SINE families. Regions of sequence similarity within the repetitive DNA elements are depicted by identical shading in the diagram. In the first step of the proposed mechanism, a Monomer family element (open rectangle 98 bp in length) is inserted into the center of a Type I Alu family element (hashed and shaded rectangle 290 bp in length) to produce a chimeric element that has two RNA polymerase III promoters (designated by the A and B boxes within each element). Transcription of the chimeric element (arrows below the element) would be initiated at a specific site upstream of each RNA polymerase III promoter region. The transcript initiated from the Monomer family promoter is drawn thicker to represent its stronger promoter activity as determined by this study. Reverse transcription of the transcript represented by the thick arrow would result in the formation of a cDNA that contains a Monomer family element attached to the right half of a Type I family element to yield the Type II element (open and shaded rectangle 270 bps in length). The Type II element would then be inserted into the genome and amplified by transcription (thick arrow below the element) and retroposition. Integration of the Type II SINE into genomic DNA would generate short direct repeats (depicted as short arrows) that flank the repetitive DNA element.

polymerase I. After the elongation reaction was completed, the DNA molecule was cleaved with *Eco*RI restriction endonuclease and denatured by boiling for three minutes. The labelled strand was then separated from the template by electrophoresis in a denaturing 7 M urea- 6% polyacrylamide gel. The single-stranded DNA probe was then eluted from the gel and concentrated by precipitation. Using this approach, we have prepared probes for both strands (using clones with inserts in opposite orientations) of a Monomer family clone (GSE 12) and Type II family clone (GAL 25). The insert from GAL 25 was digested with *Hpa*II restriction endonuclease to generate subclones containing only the left-half (GAL 25_{H2}) and right-half (GAL 25_{H10}) segments of the Type II family. The proportion of the 602 random recombinant clones that hybridized to each of the probes was determined, initially by plaque hybridization (16), and then confirmed by dot blot analysis of clones which were positive in the first screening (27). To improve the accuracy of our copy number estimate we determined the size of the inserts for twenty random clones isolated from the galago M13mp8 library and found the insert size to average 500 bps.

RESULTS

The Monomer family is an independent SINE family in the galago genome

In our original description of the galago Type II family (8), we found one example of a repetitive element (GAL 39) which apparently contained only the left half of the Type II sequence. It was not clear whether GAL 39 represented an independent SINE family or was simply a truncated Type II family clone. Additional sequence analysis of the clones generated in our initial study of galago SINE families (7,8), revealed that several left-half Type II clones (GAL 2,32,38,39) existed independently of the Type II right-half sequence (Figure 2). These monomeric Type II left-half repetitive elements were described as the galago Monomer (or Type III) family and were found to have similar sequences to methionine tRNA genes (9). Since we had already thoroughly characterized a large number of individual Alu (Type I) and Type II SINE family members in our previous studies (7,8), we decided to extend our analysis of galago SINE families by characterizing additional Monomer family members and determining their relationship to the other SINEs. To accomplish this goal, we cloned *Rsa*I digested *Galago senegalensis* genomic DNA into the M13mp8 vector to generate a random genomic library and screened the library with a Monomer family probe (GAL 39) which detected both Monomer and Type II family members. After a second more stringent screening of 60 positive clones was conducted by dot blot analysis, twelve clones (GSE 7,9,12,18,20,32,36,37,40,41,43,55) were selected for further characterization by sequence analysis.

As shown in Figure 2, the aligned sequences of 17 Monomer elements [4 clones isolated from a *Galago crassicaudatus* library (GAL, Ref. 7,8) and 12 clones isolated from a *Galago senegalensis* library (GSE)] generates a 100 base consensus sequence which demonstrates that these repetitive sequences can be distinguished as an independent SINE family. Apparently, there is no obvious difference between the Monomer family elements in these two closely related galago species as determined by our sequence analysis, nor would we expect there to be a difference. As we have observed previously for the galago Alu (Type I) and Type II SINE families (7,8), the Monomer SINEs are not completely identical in their sequences. The individual Monomer family elements show a 12% divergence on average from the consensus sequence (the most common nucleotide at each position) with only a small number of insertions and deletions occurring per clone. The divergence from the Monomer consensus sequence for individual clones is: 7.1%, 10%, 14.1%, and 10% for the GAL clones in numerical order and 25.4%, 7.1%, 11.1%, 10%, 17.5%, 10.7%, 6.1%, 16.2%, 33.6%, 5.1%, 9.1%, 13.3%, and 12.9% for the GSE clones in numerical order (Figure 2). Thus, the range of divergence is from 5.1% to 33.6% for individual Monomer family clones. A significant amount of the sequence divergence from the Monomer consensus sequence can be attributed to 4 CpG dinucleotides located at positions 4, 31, 47, and 74 in the consensus sequence (Figure 2).

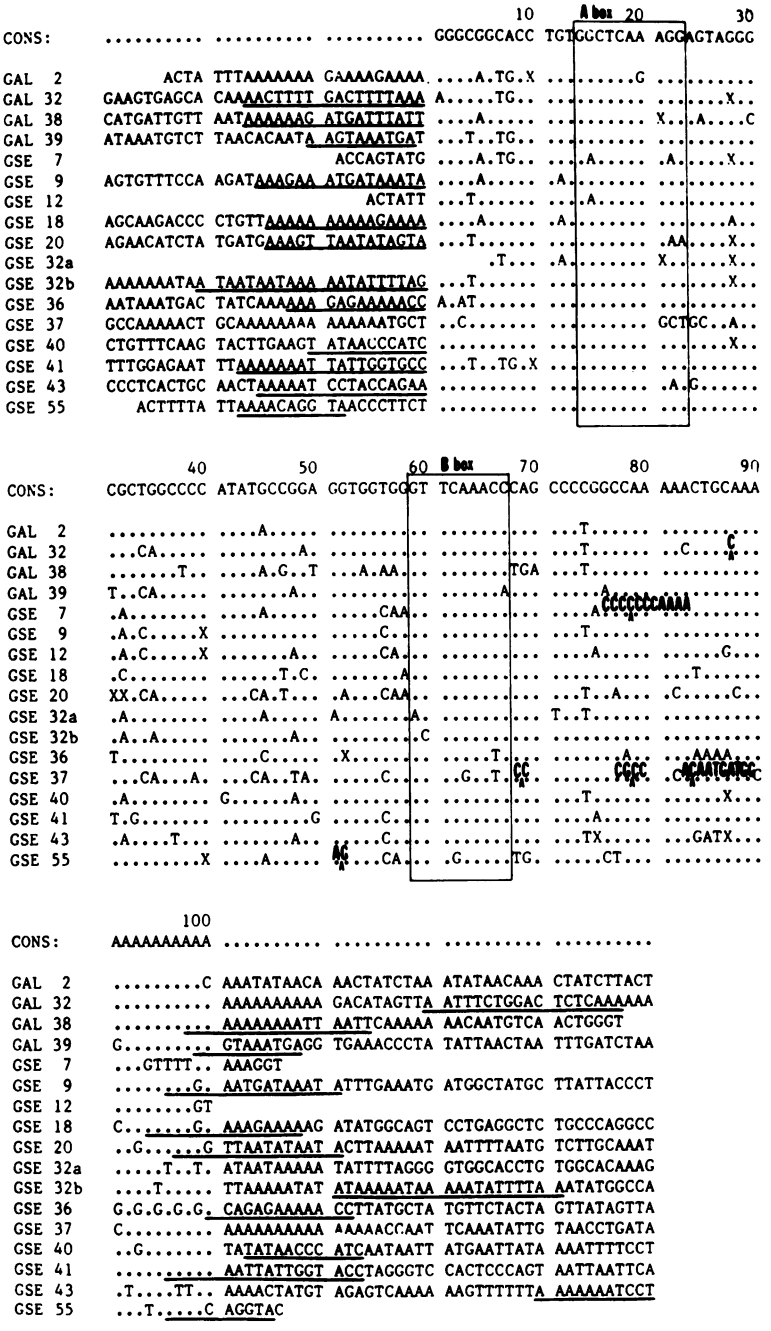

Figure 2. Consensus sequence for the galago Monomer family. The nucleotide sequence of 16 cloned Monomer family members isolated from the genomes of *Galago crassicaudatus* (GAL 2,32,38,39) and *Galago senegalensis* (GSE 7,9,12,18,20,32,36,37,40,41,43,55) are aligned to determine the most frequent base at each position (1 – 100). Dots located within the sequence of individual Monomer family clones indicate agreement with the consensus sequence (CONS). The appropriate base is given in the sequence at positions that do not agree with the consensus sequence. An x represents a deletion relative to the consensus sequence and insertions are depicted as bold lettered sequences above an arrowhead indicating their position within the sequence. Some additional sequences are included for each Monomer family clone to identify the positions of direct repeats (underlined sequences) that flank each repetitive sequence. Note that GSE 32a and GSE 32b were arranged as a tandem dimer of Monomer family elements in a single clone. The RNA polymerase III promoter blocks (A and B boxes) are outlined. Numbering begins at the first base of the consensus sequence.

Because the left-half segment of the Type II family was discovered to be related to the Monomer family, we compared the consensus sequences for these galago SINE families and found that they had a 75% sequence identity over an 80 base region (data not shown). Additional analysis demonstrates that the galago Monomer family (Figure 2) shares many common features with other primate SINEs including: 1) a precisely defined 5′ terminus (position 1 in the consensus sequence); 2) a variable oligo(dA)-rich 3′ terminus (dA-rich sequences near position 100); 3) direct repeats that flank the consensus sequence (underlined sequences); and 4) a split intragenic RNA polymerase III promoter (blocks A and B). These structural features indicate that the Monomer elements are an independent SINE family that has been dispersed throughout the galago genome by retroposition.

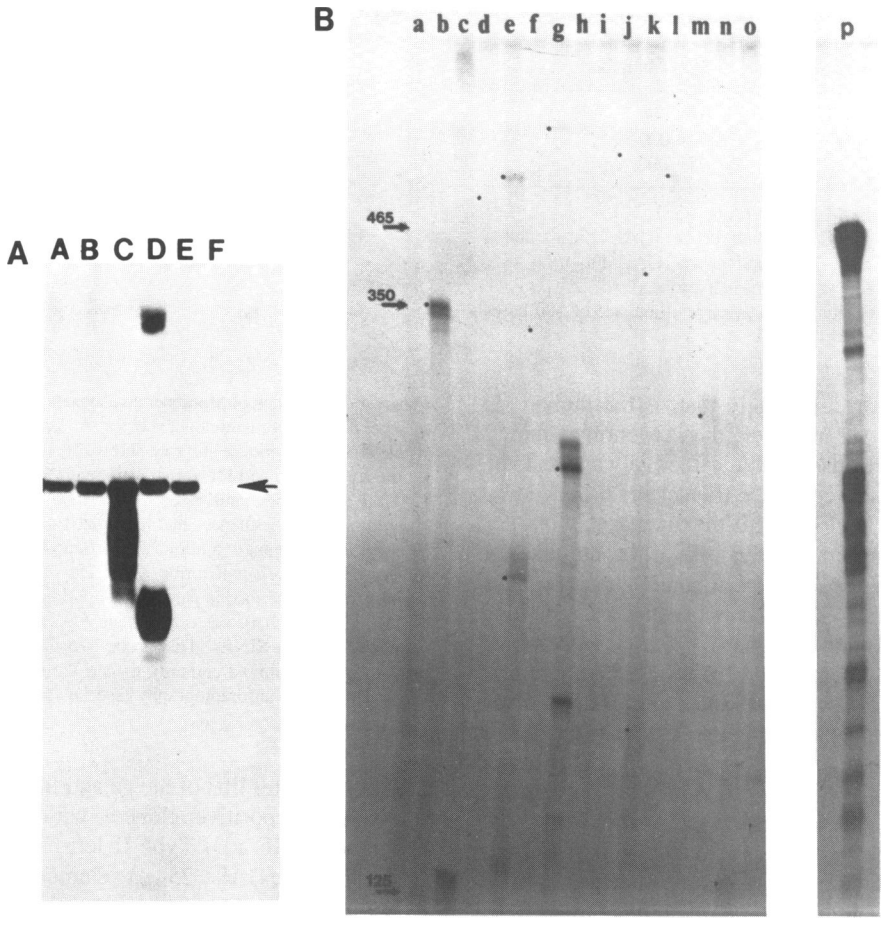

Figure 3. Transcriptional analysis of galago SINE family clones. A. The following M13mp8 clones were used as templates in an *in vitro* RNA polymerase III transcription system: A) JD 7 (a human Alu family M13mp8 clone that produces a 350 base transcript), B) GAL 19 (a galago Alu family clone) C) GAL 20 (a galago Type II family clone), D) GAL 34 (a galago Type II family clone), E) GAL 35 (a galago Type II family clone), and F) M13mp8 (control template). Each transcriptional reaction mixture contained 500 ng of template DNA and 500 ng of either M13mp8 (lanes A and F) or JD 7 (lanes B, C, D, and E) to bring the final DNA concentration to 1 μg. The arrow indicates the position of the JD 7 transcript (350 bases) which serves as an internal standard for this transcriptional study. Additional transcripts were not observed in lanes B and E. B. This panel shows the transcripts of predominately Monomer family clones. The following M13mp8 clones were used as templates in an *in vitro* RNA polymerase III transcription system: a) M13mp8 (control template), b) JD 7 (human Alu family clone which produces a 350 base transcript), c) pXbs I (a *Xenopus borealis* 5S gene clone which produces a transcript of 125 bases), d) GSE 7, e) GSE 9, f) GSE 12, g) GSE 18, h) GSE 20, i) GSE 32, j) GSE 36, k) GSE 37, l) GSE 40, m) GSE 41, n) GSE 43, o) GSE 55, and p) GAL 25 (a Type II family clone which produces a 465 base transcript). The reaction mixture contained 500 ng of template DNA and 500 ng of M13mp8 for each transcriptional assay. Lanes d through o show the transcripts produced by the Monomer clones and lane p shows the transcripts produced by the Type II family clone (GAL25) which has the strongest RNA polymerase III promoter of all clones tested in our studies. Dots mark the positions of the most prominent transcripts in each lane.

Transcriptional analysis of galago SINE family clones

Since retroposition of SINEs requires transcription, we were interested in characterizing the transcriptional properties of the three galago SINE families to learn more about the amplification of these repetitive elements. We also noted that the galago Monomer and Type II family members had an RNA polymerase III promoter apparently derived from a methionine tRNA gene (9), whereas the galago Alu (Type I) family had a promoter derived from a 7SL RNA gene (28) and wished to determine if these promoters were transcribed to the same extent. To accomplish this aim, we purified DNA from several different SINE family clones and conducted *in vitro* transcriptional studies using HeLa cell S100 extracts as our source of RNA polymerase III. As shown in Figure 3A, we directly compared the transcriptional activity of several cloned galago Type I (GAL 19) and Type II (GAL 20, 34, and 35) SINE family members with a human Alu family control clone (pJD 7) by incubating the clones together in the same transcriptional reaction mixture and separating the transcripts. The control clone (pJD 7) in this analysis represents a human Alu family element which was determined to have the strongest promoter out of 20 random Alu family clones tested (J. Randall and P. Deininger, unpublished). Densitometric analysis of the transcripts shown in Figure 3A indicates that transcription of the human Alu family clone, JD 7 (see transcripts designated by the arrow), was surpassed approximately 5-fold by the transcripts produced from the galago Type II clones, GAL 20 and GAL 34. This analysis suggests that the Type II family RNA polymerase III promoter (in clones GAL 20 and GAL 34) is transcribed more efficiently than is the human Alu family promoter in clone JD 7.

As demonstrated in Figure 3B, *in vitro* transcriptional analysis of the Monomer family clones indicates that most of these clones have an active RNA polymerase III promoter. The most highly transcribed Monomer family clones (GSE 9 and GSE 18 in lanes e and g of Fig. 3B, respectively) appear to have promoters that

Table 1. Estimated genomic copy number of the galago SINE families

SINE Family	Clones/602 screened	Copy Number[d]
Monomer	20[a]	199,000
Type II	32[b]	319,000
Type I (Alu)	26[c]	259,000

[a]Clones that hybridized only to the Monomer probe (GSE 12).
[b]Clones that hybridized to a Type II right-half probe (GAL 25_{H10}) and a Monomer probe (GSE 12) or a Type II left-half probe (GAL 25_{H2}).
[c]Clones that hybridized only to a Type II right-half probe (GAL 25_{H10}).
[d]The copy number was estimated assuming an average clone size of 500 bases and a haploid genome size of 3×10^9 bp.

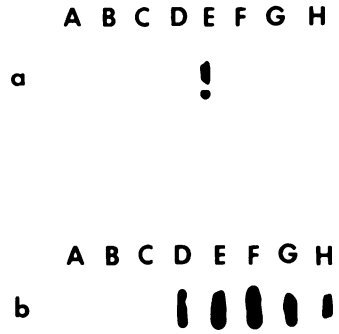

Figure 4. Distribution of Monomer and Type II right-half sequences in the genomes of several different mammalian species. A. Dot blots of genomic DNA (500 ng) from: A) cow, B) mouse, C) tree shrew, D) lemur, E) galago, F) owl monkey, G) rhesus monkey, and H) human were hybridized to the Monomer family probe, GAL 2, and the abundance of related sequences was determined by autoradiography. Sequences that hybridized to the Monomer family probe were detected only in galago genomic DNA (lane E). The same result was obtained with the Type II left-half probe, GAL 25_{H2} (data not shown). B. An identical dot blot to that described in part A was hybridized with a Type II right-half probe (GAL 25_{H10}) which also detects primate Alu family elements due to the sequence similarity of these SINEs. This probe hybridized to the genomic DNAs of all primate species (almost certainly to Alu family sequences within the genome) as determined by autoradiography (see lanes D, E, F, G, and H).

are transcribed slightly more efficiently than is the human JD 7 Alu family clone (lane b). Moreover, these Monomer family clones do not appear to be transcribed as effectively as the Type II family clones as demonstrated by the transcripts produced by clone GAL 25, the Type II clone with the strongest promoter of all clones tested (see lane p Fig. 3B). Like the human Alu family (23), both the Monomer and Type II family clones produce transcripts of variable lengths. This occurs because the repetitive element itself does not contain a signal for the termination of transcription by RNA polymerase III. Instead, transcription is terminated in most cases in the 3′ flanking region at fortuitous termination sites located in the cloning vector. Some of the clones shown in Figure 3 generate multiple transcripts of different lengths indicating that incomplete termination occurs at various points in the sequence. We have mapped the positions of these termination sites in the vector sequence (data not shown) and, as expected, find a RNA polymerase III terminator sequence (four or more T residues) at the site of termination in most cases (29). This indicates that some of the longest transcripts observed in Figure 3 result from transcriptional read-through at the first termination site encountered.

To determine the initiation site for transcription of the galago SINE family clones, we carried out S1 nuclease mapping analysis on transcripts produced by *in vitro* transcription for two Monomer (GAL 2 and GAL 39) and Type II (GAL 5 and GAL 25) family clones. Our results demonstrated that the transcriptional initiation site for all the Monomer and Type II family clones tested was the expected 5′ end of the repetitive DNA sequence which corresponds to the first base of the consensus sequence for each SINE family (data not shown). This analysis also confirmed our previous assumption that terminations were occurring at various sites in the downstream flanking region of the galago SINE family clones.

Copy number and species-specificity of the galago SINE families

Because the galago Type II family members cross-hybridize with probes for either Alu (Type I) or Monomer family members, direct quantitation of the three galago SINE families by hybridization techniques is difficult. To overcome this problem, we have chosen to use library screening procedures (16) to estimate the copy numbers of the galago SINE families. For this analysis, we have used our size-selected *Galago senegalensis* M13mp8 library, which has inserts averaging about 500 base pairs in length. Thus, the probability that individual clones from this library would contain more than one galago repetitive DNA element is fairly low. Assuming that each positive clone contains only one repetitive DNA element, we discovered that we could differentiate between the three galago SINE families by hybridizing filter lifts of the galago library and dot blots (second screening of positive clones) with separate probes for the Monomer (GSE 12), Type II left-half (GAL 25_{H2}), and Type II right-half (GAL 25_{H10}) elements. Thus, clones which hybridized to the Monomer and/or Type II left-half probes, but not to the Type II right-half probe were designated as Monomer family clones. Clones which hybridized to the Type II right-half probe, but not to the Monomer and/or Type II left-half probes were counted as Type I family clones and clones which hybridized to all three SINE probes were considered to be Type II family clones. Using this method to estimate the number of galago SINE family members, we screened 602 random clones from the galago library and discovered that 20 clones contained Monomer elements, 32 clones contained Type II elements, and 26 clones contained Type I elements. As shown in Table 1, we estimate from these data that the three major galago SINE families are present in the range of approximately 200,000 to 320,000 copies per family in the galago genome.

The abundance and similarity in copy numbers for the Monomer, Type II, and Type I (Alu) families in the galago genome suggested that all three SINE families might have originated very early in primate evolution. Since the 300 base pair dimeric form of the human Alu family had been found in all primates, but not in rodents (2), we wished to determine the distribution of the Monomer and Type II SINE families in the genomes of other primate and mammalian species. This was accomplished by hybridizing dot blots of genomic DNA from different mammalian species with probes specific for the Monomer (GAL 2), Type II left-half (GAL 25_{H2}), and Type II right-half (GAL 25_{H10}) sequences as described in Figure 4. This analysis of mammalian genomic DNA demonstrates that the Monomer and Type II SINE families are highly amplified only in the galago genome (Figure 4a), whereas the Alu family (as detected by the Type II right-half sequence probe) is present in high copy numbers in the genomes of all primates tested including lemur, galago, owl monkey, rhesus monkey, and human (Figure

	A BOX	B BOX
tRNA Pol III CONSENSUS	R R Y N N A R Y G G . 33-43 bp.	G T T C R A N N C
MET tRNA	G G C G C A G C G G . .33 bp.	.G A T C G A A A C
Type II CONSENSUS	A G C A C A G T G G . .34 bp.	.G T T C G A A C C
GAL 25	. . . . .	. . . . .
Monomer CONSENSUS	G G C T C A A a G G . .35 bp.	.G T T C A A A C C
GAL 32	. . . . .	. . . . .
GSE 9	. . . . .	. . . . .
GSE 18	. . . . .	. . . . .
GSE 40	. . . . .	. . . . .
GSE 41	. . . . .	. . . . .
Type I CONSENSUS	G G C A T g G T G G . .61 bp.	.G T T t G A G A C

Figure 5. Comparison of the RNA polymerase III promoters of the galago SINE families. The tRNA RNA polymerase III promoter consensus sequence is presented along with the equivalent promoter regions from a human methionine (MET) tRNA gene, the Type II family consensus, the Monomer consensus, and the Type I family consensus sequence. The promoter regions for several Type II (GAL25) and Monomer (GAL 32, GSE 9, GSE 18, GSE 40, GSE 41) family clones are presented as dots to indicate their agreement with their respective consensus sequence promoters. For the tRNA promoter consensus sequence: R designates a purine nucleotide, Y designates a pyrimidine nucleotide, and N represents any nucleotide commonly found in DNA. Deviations from the RNA polymerase III promoter consensus sequence are indicated by a small letter at the appropriate position. The distance between the A and B boxes for each promoter is also indicated.

4b). This indicates that the Monomer and Type II families were formed and apparently amplified since the divergence of the prosimian species, galago and lemur, which occurred about 45 million years ago. These results agree with sequence divergence data which suggests that most mammalian SINE families are homogeneous and species-specific in their sequences due to their recent amplification on an evolutionary time scale (2).

DISCUSSION

In addition to the Alu family found in the genomes of humans and other primates, the galago genome has two other major SINE families. These three galago SINE families have been described previously as the Type I or Alu (7), Type II (8), and Monomer (Type III) families (9). Almost all members of these galago SINE families share common features with other SINEs including: 1) a precisely defined 5′ terminus; 2) a variable oligo(dA)-rich 3′ terminus; 3) a split intragenic RNA polymerase III promoter; and 4) direct repeats that flank the repetitive element. By these criteria, we have demonstrated in this study that the Monomer elements shown in Figure 2 are an independent SINE family in the galago genome and are not the truncated left-half regions of several Type II elements. The fact that the Monomer elements are an independent SINE family capable of retroposition suggests a mechanism for the formation of the chimeric Type II family (Figure 1). Based on the sequence analysis of several subfamilies, we have proposed that the galago Type II family resulted from the integration of a Monomer family element into the center of a Type I Alu family member (10). This specific integration event apparently allowed the Type II family to spread itself by retroposition throughout the galago genome using the RNA polymerase III promoter of the Monomer element.

Since our first analysis of the major galago SINE families suggested that all three were highly copied in the genome (7,8), we attempted to determine their distribution in the genomes of other primate species to learn more about the origin of each SINE family. Dot blot analysis of genomic mammalian DNA indicated that the Monomer and Type II families were highly amplified only in the galago genome (Figure 4a), whereas the Alu (Type I) family was present in high copy number in the genomes of all primates tested (Figure 4b). This analysis demonstrated that the Monomer and Type II families were unique to the galago genome indicating that these two SINE families were apparently formed and amplified much more recently on an evolutionary time scale than was the Alu family (2). Moreover, we believe that the Monomer family must have originated prior to the Type II family as evidenced both by its higher divergence from the consensus sequence (12% for the Monomer family versus 6% and 9% for the two subfamilies of the Type II; Ref. 2), as well as the observation that the Type II family was almost certainly derived from the fusion of a Monomer element with an Alu (Type I) family member (10).

Copy number analysis for the three galago SINE families was carried out to determine the relative abundance of each family and the rate of SINE family amplification in the galago genome. Our analysis demonstrated that the Alu family is present at a lower copy number in the galago genome than it is in the human genome (259,000 in galago versus about 500,000 in human; Ref. 6). This may have occurred due to the presence of the two other actively amplifying SINE families in galago which may have competed with Alu family elements for the retroposition machinery and available target sites resulting in an early saturation of the galago genome with SINE sequences. It is also possible that mutations occurred in the galago Alu family early in its evolution, which merely made it amplify less effectively than the Alu family in other primates. This is supported in part by our previous studies which demonstrated that the galago Alu (Type I) family has species-specific differences in its sequence compared to the Alu family in the other primates (7,30). Surprisingly, the most recently amplified SINE families in the galago genome (the Monomer and Type II families) were found to be present in similar copy number to the galago Alu (Type I) family (see Table I). In fact, the most recently formed galago SINE family (the Type II) is also the SINE family with the highest copy number in the galago genome. Comparison of sequence divergence within these SINE families suggests that the Alu family may be as much as twice as old as the Type II family and appreciably older than the Monomer family (2). What has allowed the Monomer, and especially the Type II families to amplify more effectively than the Type I (Alu) family in the galago genome? Did the galago Alu family simply stop amplifying and allow the other SINE families to catch up or do the Monomer and Type II families have some advantage in the retroposition process?

Our analysis of the *in vitro* transcriptional data (Figure 3) suggests that the overall order of transcriptional activity for the galago SINE families is Type II family > Monomer family > Alu family. Thus, one factor that may have affected the rate of amplification of the Monomer and Type II families is the acquisition of a RNA polymerase III promoter that is transcribed more efficiently than is the Alu family promoter. This is certainly a likely possibility for any repetitive element that undergoes an RNA-mediated transposition as the importance of the promoter to retroposition is supported by studies of the yeast TY I element (31). Transcriptional analysis of numerous SINE family members indicates that the Monomer and Type II family clones are transcribed 2 to 5 fold more effectively than is our most efficiently transcribed human Alu family clone, JD 7 (Figure 3). These differences in transcriptional activity probably result from differences in the RNA polymerase III promoter found within each SINE family.

Since there are notable differences in the sequence and transcriptional activity of individual SINE family members, it does not seem likely that all members of a SINE family are actually involved in the retroposition process (2). The fact that some SINEs, such as the galago Type II family, can be divided into subfamilies (10), suggests that a small group of master sequences has been responsible for the amplification of most SINE family members. This is also supported by evolutionary analysis of human Alu family members which suggests that amplification of this SINE family has been dominated by one, or just a few, master gene(s) (32,33). The consensus sequence for a SINE family actually represents our best estimate of the most actively retroposing or master sequence (1). Thus, individual SINEs that show a high degree of sequence identity to the consensus are the SINEs that most accurately reflect the properties of the master sequence. Analysis of the Monomer family (Figure 2) and Type II family (8) sequences shows that only a limited number of clones actually match the consensus sequence within the promoter region. These clones are GAL 25 for the Type II family and clones GAL 32, GSE 9, GSE 18, GSE 40, and GSE 41 for the Monomer family (Figure 5) and are our best examples of individual SINEs that should reflect the true promoter strengths of the parental sequence for these SINE families. As predicted, these galago clones were some of the most highly transcribed by RNA polymerase III demonstrating a greater promoter strength than most of the other clones tested, including the human Alu family clone, JD7.

Additional analysis of these RNA polymerase III promoter sequences points out several other interesting features. Both the Monomer and Type II family promoters originally were derived from a methionine tRNA gene (9). The sequence of that promoter is presented in Figure 5 along with the consensus sequence for all three galago SINE family (7–9) and all eukaryotic tRNA promoters (34). The Type I family consensus promoter differs considerably from the Monomer and Type II consensus promoters in that the distance between the A and B promoter blocks is 61 bp for the Type I versus about 34 bp for the tRNA-like promoters. Numerous studies have indicated that the relative positions of the A and B blocks with respect to each other affects the efficiency of transcription by RNA polymerase III (35,36). Moreover, the A and B blocks of the Type I family promoter contain nucleotides that do not match the RNA polymerase III promoter consensus sequence (Figure 5) which may affect promoter strength. The Monomer family consensus promoter also has one mismatch to the RNA polymerase III promoter consensus sequence. However, the Type II consensus promoter is perfectly matched to the RNA polymerase III promoter consensus sequence and actually represents the most efficiently transcribed promoter in our studies. Finally, as shown in Figure 5, the consensus sequence for the promoter blocks differs for each galago SINE family. Whether these changes result in the formation of a stronger promoter for rapidly amplifying SINE families or are instead adaptations of the SINE family to allow the promoter to function better in a more varied genomic environment will require further analysis.

EMBL ACCESSION NUMBERS

The EMBL accession numbers for the galago Monomer, Type II, and Type I SINE family sequences are: GAL 2 (X55906), GAL 5 (X00097), GAL 19 (X00116), GAL 20 (X00108), GAL 25 (X00100), GAL 32 (X03322), GAL 34 (X00102), GAL 35 (X00111), GAL 38 (X03327), GAL 39 (X03328), GSE 7 (X55920), GSE 9 (X03329), GSE 12 (X55910), GSE 18 (X03331), GSE 20 (X03332), GSE 32 (X03333), GSE 36 (X03334), GSE 37 (X55915), GSE 40 (X03335), GSE 41 (X03336), GSE 43 (X03337), GSE 55 (X03338).

ACKNOWLEDGEMENTS

We would like to thank A. Jennifer Randall for providing the human Alu family clone, pJD 7, for these studies and Dr. Sal Caradonna for providing us with HeLa cells. This work was supported in part by USPHS grant HG00340.

REFERENCES

1. Rogers, J. (1985) Int. Rev. Cytol. **93**, 187–279.
2. Deininger, P.L. and Daniels, G.R. (1986) Trends in Genetics **2**, 76–80.
3. Weiner, A.M., Deininger, P.L., and Efstratiadis, A. (1986) Ann. Rev. Biochem. **55**, 631–661.
4. Deininger, P.L. (1989) In Howe, M. and Berg, D. (eds.), Mobile DNA. ASM Press, Washington D.C., pp. 619–636.
5. Rinehart, F.P., Ritch, T.G., Deininger, P.L., and Schmid, C.W. (1981) Biochemistry **20**, 3003–3010.
6. Deininger, P.L., Jolly, D.J., Rubin, C.M., Friedmann, T., and Schmid, C.W. (1981) J. Mol. Biol. **151**, 17–33.
7. Daniels, G.R., Fox, G.M., Loewensteiner, D., Schmid, C.W., and Deininger, P.L. (1983) Nucleic Acids Res. **11**, 7579–7593.
8. Daniels, G.R. and Deininger, P.L. (1983) Nucleic Acids Res. **11**, 7595–7610.
9. Daniels, G.R. and Deininger, P.L. (1985a) Nature **317**, 819–822.
10. Daniels, G.R. and Deininger, P.L. (1985b) Nucleic Acids Res. **13**, 8939–8954.
11. Rogers, J. (1983) Nature **301**, 460.
12. Jagadeeswaran, P., Forget, B.G., and Weissman, S.M. (1981) Cell **26**, 141–142.
13. Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Manser, T., and Gesteland, R.F. (1981) Cell **26**, 11–17.
14. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press, Cold Spring Harbor.
15. Messing, J., Crea, R., and Seeburg, P.H. (1981) Nucleic Acids Res. **9**, 309–321.
16. Benton, W.D. and Davis, R.W. (1977) Science **196**, 180–182.
17. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A., and Roe, B. (1980) J. Mol. Biol. **143**, 161–178.
18. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
19. Biggen, M.D., Gibson, T.J., and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA **80**, 3963–3965.
20. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. **7**, 1513–1523.
21. Messing, J. (1983) Methods Enzymol. **101**, 20–78.
22. Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983) Nucleic Acids Res. **11**, 1475–1489.
23. Fuhrman, S.A., Deininger, P.L., LaPorte, P., Friedmann, T., and Geiduschek, E.P. (1981) Nucleic Acids Res. **9**, 6439–6456.
24. Manley, J.L., Fire, A., Cano, A., Sharp, P.A., and Gefter, M.L. (1980) Proc. Natl. Acad. Sci. USA **77**, 3855–3859.
25. Chaconas, G. and van de Sande, J.H. (1980) Methods Enzymol. **65**, 75–85.
26. Berk, A.J. and Sharp, P.A. (1977) Cell **12**, 721–732.
27. Traina, V. and Cohen, J.C. (1983) BRL Focus **5**, 5–6.
28. Ullu, E. and Tschudi, C. (1984) Nature **312**, 171–172.
29. Bogenhagen, D.F. and Brown, D.D. (1981) Cell **24**, 261–270.
30. Slagel, V., Flemington, E., Traina-Dorge, V., Bradshaw, Jr., H. and Deininger, P.L. (1987) Mol. Biol. Evol. **4**, 19–29.
31. Garfinkel, D.J., Boeke, J.D., and Fink, G.R. (1985) Cell **42**, 507–517.
32. Deininger, P.L. and Slagel, V.K. (1988) Mol. Cell. Biol. **8**, 4566–4569.
33. Britten, R.J., Stout, D.B., and Davidson, E.H. (1989) Proc. Natl. Acad. Sci. USA **86**, 3718–3722.
34. Ciliberto, G., Raugei, G., Costanzo, F., Dente, L., and Cortese, R. (1983) Cell **32**, 725–733.
35. Galli, G., Hofstetter, H., and Birnstiel, M.L. (1981) Nature **294**, 626–631.
36. Ciliberto, G., Traboni, C., and Cortese, R. (1982) Proc. Natl. Acad. Sci. USA **79**, 1921–1925.