

1-1-1996

## The mobile genetic element Alu in the human genome

Gabriel E. Novick  
*Florida International University*

Mark A. Batzer  
*University Medical Center New Orleans*

Prescott L. Deininger  
*University Medical Center New Orleans*

Rene J. Herrera  
*Florida International University*

Follow this and additional works at: [https://digitalcommons.lsu.edu/biosci\\_pubs](https://digitalcommons.lsu.edu/biosci_pubs)

---

### Recommended Citation

Novick, G., Batzer, M., Deininger, P., & Herrera, R. (1996). The mobile genetic element Alu in the human genome. *BioScience*, 46 (1), 32-41. <https://doi.org/10.2307/1312653>

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

# The Mobile Genetic Element *Alu* in the Human Genome

*Transposable elements represent a dramatic source of genetic variation*

Gabriel E. Novick, Mark A. Batzer, Prescott L. Deininger, and Rene J. Herrera

Genetic material has been traditionally envisioned as relatively static with the exception of occasional, often deleterious mutations. The sequence DNA-to-RNA-to-protein represented for many years the central dogma relating gene structure and function. Recently, the field of molecular genetics has provided revolutionary information on the dynamic role of repetitive elements in the function of the genetic material and the evolution of humans and other organisms.

*Alu* sequences represent the largest family of short interspersed repetitive elements (SINES) in humans, being present in an excess of 500,000 copies per haploid genome. *Alu* elements, as well as the other repetitive elements, were once considered to be useless. Today, the biology of *Alu* transposable elements is being widely examined in order to deter-

---

*Alu* elements,  
responsible for  
innumerable genetic  
disorders, can be  
markers for human  
identification and are  
raw material  
for evolution

---

mine the molecular basis of a growing number of identified diseases and to provide new directions in genome mapping and biomedical research.

Only a small fraction of the human genome consists of functional, or protein-encoding, genes. If all the nucleotide information were translated,  $10^9$  amino acids, or approximately 3 million average-size proteins, would be encoded. Based on the fully sequenced genomes of simple organisms and considerations of the higher complexity of humans, it is estimated that no more than 60,000 to 100,000 proteins, in addition to functional RNAs, are essential for the survival of the species (Alberts et al. 1989) and as a result are conserved throughout evolution. Considering that many genes are repetitive, only approximately 10%–20% of the human DNA encodes functional sequences. What is the function of the remaining DNA?

## Discovery and classification of repetitive elements

When the DNA double helix of higher organisms (species other than bacteria and viruses) is split open or denatured into two single strands by heat and left to spontaneously reanneal, the entire DNA is not likely to become double-stranded at the same rate. In contrast, all parts of bacterial and viral DNA reanneal at approximately the same rate. This phenomenon was observed in the now classical experiments of Britten and Kohne (1968) and was correctly interpreted as an indication of greater sequence complexity among eukaryotes.

These experiments distinguish three populations of eukaryotic DNA based on the time that it takes DNA to again become double-stranded. There is a quickly reannealing component representing 25% of the total DNA, an intermediate component that represents 30% of the total DNA, and a slow component that represents 45% of the total DNA (Britten and Kohne 1968).

When the slow component was analyzed by DNA sequencing, it was found to be made up of single-copy sequences, which include the genes that encode proteins, while the fast and intermediate components reflected the presence of repetitive sequences. Repetitive DNA can find a complementary sequence on the other strand more rapidly than can single-copy sequences, because reiterate DNA can bind to any one of

---

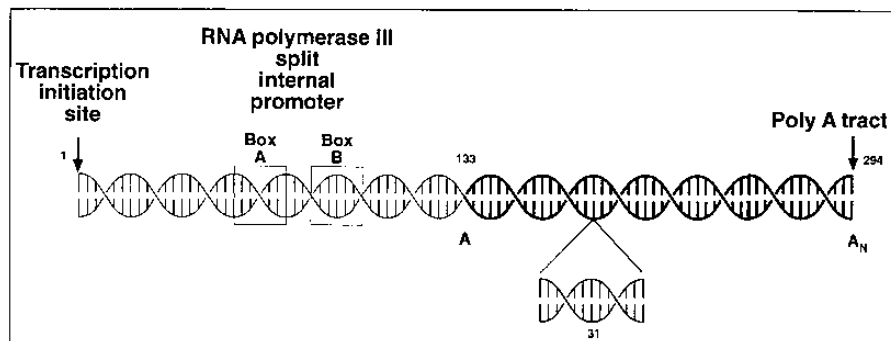
Gabriel E. Novick is a doctoral candidate and Rene J. Herrera is an associate professor in the Department of Biological Sciences, Florida International University, Miami, FL 33199. Mark A. Batzer is an assistant professor in the Pathology Department and Prescott L. Deininger is a professor in the Department of Biochemistry and Molecular Biology, Stanley S. Scott Cancer Center, Louisiana State University Medical Center, New Orleans, LA 70112. Deininger is also associate director of basic research at the Stanley S. Scott Cancer Center. Reprint requests and correspondence should be addressed to Rene J. Herrera. © 1996 American Institute of Biological Sciences.

the copies on several sites. The fast component contains small (a few nucleotides long), highly repetitive DNA sequences, which are usually found in tandem. The intermediate component contains moderately reiterated DNA sequences (Brosius 1991).

Intermediate repetitive DNA shows two common characteristics: The repetitive units, which can be up to several thousand nucleotides long, are similar to one another but not identical, and they are interspersed along the genome (not tandemly arrayed). Within this category are the transposons, elements that are able to move from one genomic location to another. This movement can be mediated by DNA or RNA. In RNA-mediated transposition, the transposon is transcribed to an RNA intermediate, this RNA is reverse transcribed into DNA, and this copy DNA (cDNA) is reinserted into the genome. The elements mobilized in this manner are known as retroposons (Rogers 1985).

Retroposons can be divided into two superfamilies. The viral superfamily represents ancient viral integration events and includes the *Copia* family in *Drosophila* (Mount et al. 1985) and *THE1* in humans (Paulson et al. 1985). The nonviral retroposon superfamily can be further divided into short and long interspersed elements (SINES and LINES). SINES are usually 75 to 500 base pairs (bp) in length, and LINES can be as long as 7000 bp. SINES contain internal RNA polymerase III promoters, while LINE transcription is under the control of the RNA polymerase II promoters that are generally present in the adjacent sequences flanking the retroposons. Also, SINES do not encode specific enzymes, while many LINES encode at least the reverse transcriptase used in their own retroposition. Being transcribed by RNA polymerase III from internal promoters allows SINES to remain transcriptionally active even after retroposition to new chromosomal sites.

The best characterized LINE is the L1 family (Rogers 1985), which is present in excess of 50,000 copies interspersed throughout each mammalian genome. The *Alu* family is the most abundant and best charac-



**Figure 1.** The dimeric structure of the *Alu* element. The two halves are linked by an adenine-rich area marked by a capital A at position 133. The right monomer (heavy lines) includes a 31-base pair insertion, and the left half contains the RNA polymerase III promoter (boxes A and B). Many *Alu* elements possess an adenine-rich region at the end, marked by an A<sub>N</sub> (at position 294 in this diagram).

terized SINE family in primates (Daniels et al. 1983).

### Distribution and structure of the *Alu* family

The *Alu* family of repetitive elements was originally identified as a fraction of renatured repetitive DNA that was distinctively cleaved with the restriction enzyme *Alu* I (Houck et al. 1979). The *Alu* family constitutes approximately 5% of the total mass of the human genome, and there are more than 500,000 copies per haploid genome (Rinehart et al. 1981).

The structure of an *Alu* element, as shown in Figure 1, presents a dimeric organization in which the right monomer (heavier lines) is 31 nucleotides longer than the left one. The left half of the element contains the RNA polymerase III promoter boxes A and B. The promoter directs transcriptional initiation to a specific site upstream of box A. Although box A, located between and including nucleotides 5 and 31, seems to determine strength and accuracy of initiation, only box B, which includes positions 70–100, appears to be essential for transcription (Fuhrman et al. 1981). *Alu* elements possess a short A-rich sequence between the two dimers and a 3' oligo A track of up to 200 bps, which are characteristic of all SINES (Economou et al. 1990). The repeat is flanked by direct repeats derived by duplication of target sequences at the site of integration.

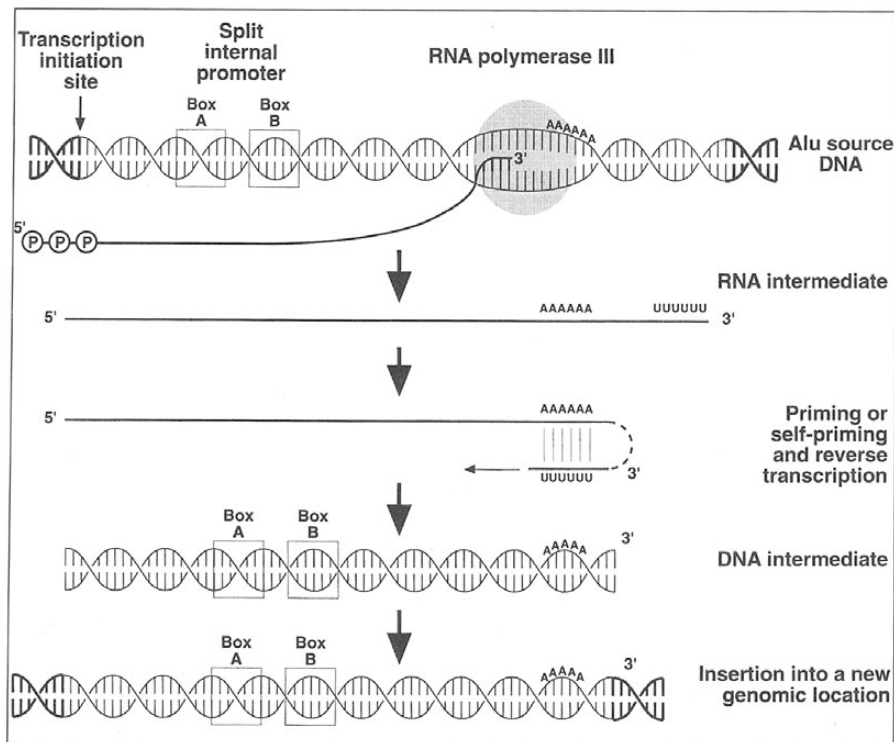
*Alu* elements are ancestrally derived from the 7SL RNA gene (Ullu

and Tschudi 1984). 7SL RNA is a small constituent of the signal recognition particle (SRP) involved in the transportation of proteins into the lumen of the endoplasmic reticulum in the cytoplasm of the cell. The SRP binds to a specific signal contained in secretory proteins and targets them to the endoplasmic reticulum (Walter and Blobel 1982). Although *Alu* elements have diverged, they retain approximately 90% sequence similarity to the 7SL RNA gene. However, the gene has 150 bp in its middle that is not found in the *Alu* family (Deininger 1989).

### Mechanism of *Alu* retroposition

The mechanism by which *Alu* and other SINES retropose consists of steps in which the element is transcribed, reverse transcribed, and then integrated into a new genomic location (Figure 2).

**Transcription.** Transcription of *Alu* elements (similar to almost all other SINES) is initiated by RNA polymerase III at a site upstream of the internal split promoter. Transcription continues along the entire element and into the flanking sequence where the polymerase terminates its reading (short poly T track). In this way, the *Alu* RNA transcript retains its promoter after transcription, but the promoter is not sufficient for effective transcription in a new location. In vivo studies show that the 7SL RNA gene requires sequences upstream of the promoter, which



**Figure 2.** The most widely accepted model of *Alu* amplification. The lighter lines represent the transposable element. An RNA intermediate transcribed by RNA polymerase III is produced. That transcript contains a poly A stretch at the 3' end of the element and a short poly U track downstream, which represents the polymerase III termination signal. The 3' end of the transcript can fold back, and the poly U pairs with the poly A track of the element and primes its own reverse transcription. Alternatively, an exogenous sequence may be involved in priming reverse transcription. The cDNA copy inserts itself into staggered nicks in a new genomic location. The filling of the resulting gaps in the host creates direct repeats (heavy lines) flanking the elements.

are not carried by the SINE during retroposition, for efficient expression (Ullu and Weiner 1985). Because these sequences are required for *Alu* transcription in vivo, the vast majority of *Alu* elements are transcriptionally silent unless inserted into a favorable genomic location that fortuitously provides them with the 5' *cis* regulatory sequences.

In addition, after insertion most SINEs are subject to methylation that is tissue and temporally specific. This methylation results in a tenfold higher than normal mutation rate at these sites (Batzer et al. 1990). Methylation, by itself, and the accumulated mutations within promoter sequences correlate with inhibition of transcription (Deininger and Batzer 1993). Although most SINEs are transcriptionally silent (Schmid and Maraia 1993), occasionally sequences in the target site of inser-

tion act in conjunction with the RNA polymerase III promoter to initiate transcription.

Some *Alu* family members inserted into genes are transcribed by polymerase II as part of nascent mRNA. *Alu* sequences located in introns are transcribed by polymerase II and then spliced out with the rest of the intron RNA before translation of the RNA into protein (such transcription is called piggyback transcription). *Alu* sequences are sometimes found in exons (translated sequences) and may become part of peptides in some cases (Makalowski et al. 1994). As part of proteins, the amino acids corresponding to *Alu* sequences provide raw genetic material for evolution.

**Reverse transcription.** The first event involved in reverse transcription is priming. It has been postulated that the uracil residues present at the 3'

end of the transcript, derived from the RNA polymerase III termination signal (Figure 2), fold back, the uracil bases pair to an internal A rich area; and the stretch of uracils then functions as a primer for reverse transcription (Jagadeeswaran et al. 1981). It is also possible that an external primer may be involved in this process. A complementary strand to the RNA transcript is copied, starting at the primer binding site, by the enzyme reverse transcriptase.

The source of this enzyme is still a matter of controversy. RNA viruses called retroviruses use their own reverse transcriptase to synthesize a cDNA intermediate before insertion into the host genome. Retroviral infections or endogenous retroviral-like sequences (Martin et al. 1981) may provide the enzyme for SINE reverse transcription. Most L1 elements (the most important LINE family in humans) are truncated; however, some of them are known to code for a functional reverse transcriptase (Mathias et al. 1991). Because mammalian genomes contain up to 50,000 copies of L1 elements, these elements constitute a prolific source of endogenous reverse transcriptase.

**Integration.** Integration of the cDNA into a new location requires nicks in the genomic DNA (one in each DNA strand) at the target site for integration. Repair of the nick after insertion results in the formation of flanking direct repeats.

The integration process is fairly random. However, there appears to be alternating distribution of SINEs and LINES along the chromosome, based on in situ hybridization studies of Giemsa-stained metaphase chromosomes (Korenberg and Rykowski 1988). The target sites for SINE integration tend to be A+T rich regions (Batzer et al. 1990). Because SINEs are highly G+C rich elements, their insertion creates small G+C islands in A+T rich regions.

## Origins and evolution of *Alu* elements

Although it is generally accepted that the *Alu* family originated from a modified RNA transcript of the

7SL RNA gene (Ullu and Tschudi 1984), two models have been proposed to explain the expansion of the repetitive element family. These models are called the transposon model and the master gene model.

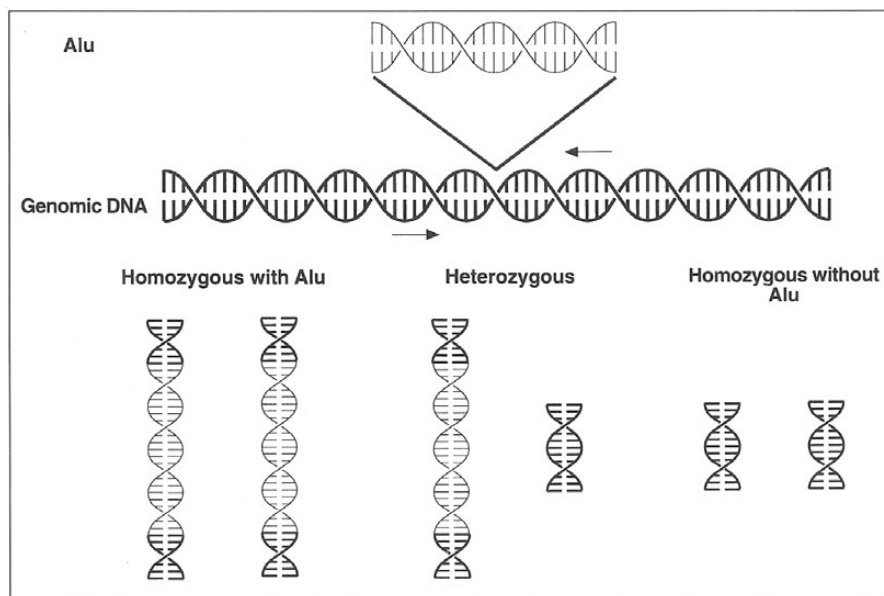
The transposon model postulates that many SINEs generate new elements, each with the same chance of being transpositionally active. This mechanism would generate sequences that diverge from the original sequence as the multiplication process continues. The nucleotide differences among *Alus* would include every possible mutation along the entire length of the elements. The sequences of the members of the family would diverge independently as different members multiply and accumulate mutations.

Data shows that the amplification rate (number of insertions as a function of time) is not exponential as would be expected if the transposon model is correct and other factors (e.g., availability of reverse transcriptase) were not limiting. In fact, *Alu* amplification rate is highly variable, with periods of high and low amplification rates. The recent *Alu* amplification rate, estimated at one insertion fixed in the population per 5000 years, is approximately 100-fold slower than at its peak approximately 40–50 million years ago (Shen et al. 1991).

The observation of a few clearly defined subfamilies of elements also does not support the transposon model. Subfamilies are defined according to the presence of certain characteristic (diagnostic) mutations. For example, a given *Alu* subfamily may be characterized by the presence of a G, instead of the more typical T, at nucleotide position 156.

The master gene model (Deininger et al. 1992) proposes that most SINEs are derived from one or a few active loci. This model predicts a linear amplification rate controlled by the amplification of the master genes. Mutations in the master genes would generate new subfamilies and might cause variations in the amplification rate (Deininger and Batzer 1993).

For example, at time zero an unmutated master gene would generate a family of elements with a sequence identical to itself. Over



**Figure 3.** *Alu* polymorphism PCR assay. Specific PCR primers (arrows) directed to the single copy sequences that flank the *Alu* insertion (light lines) are used for amplification. In individuals homozygous for a particular insertion, only one large DNA fragment containing the *Alu* insertion is likely to be amplified (representing the same locus on homologous chromosomes). In individuals homozygous for the lack of the *Alu* insertion, only one smaller DNA fragment is likely to be generated. Heterozygous individuals are likely to yield two amplification products, one made up of small DNA fragments and the other of larger pieces of DNA, representing the presence of the insertion in only one of the two homologues.

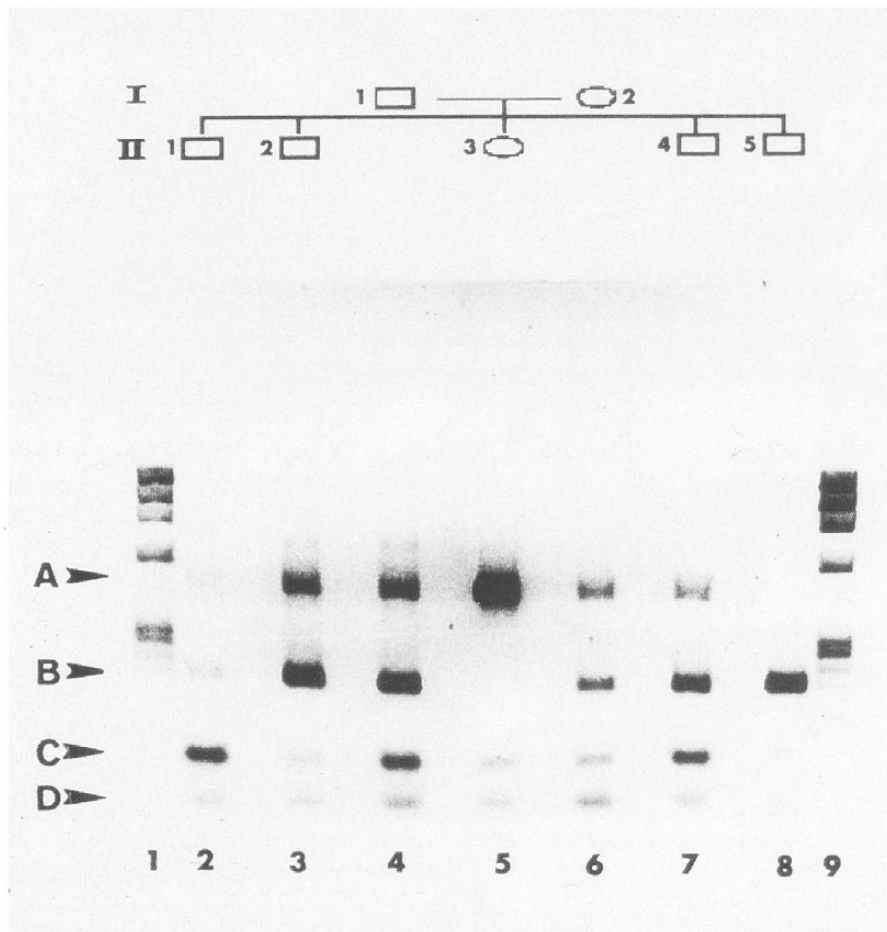
time, this master gene would mutate, and subsequent copies of the sequence would constitute a subfamily, which is slightly different from the original group. Each new mutation in this master gene would not only be added to the previous one and generate a new subfamily, but it might also affect the rate of amplification of the family.

In the case of *Alu* elements, subfamilies can be arrayed in a sequential manner (Britten 1994, Britten et al. 1989, Shen et al. 1991) in which the older the subfamily, in evolutionary terms, the more random variations are observed within the subfamily members. For example, the number of differences present among the older primate-specific (PS) subfamily members is greater than the number observed among a human-specific (HS) subfamily.

A recently inserted human specific (HS-1) subfamily was identified in humans with only a few copies located in chimp and gorilla genomes (reviewed in Batzer et al. 1995). Another human specific subfamily (HS-2) was exclusively found in humans. In addition, a third subfam-

ily, the Sb-2, has human-specific members. Some of the members of HS-1, HS-2, and Sb-2 subfamilies have inserted so recently that not all humans have the insertion in a particular locus. This observation represents the basis for the use of polymorphic *Alu* insertions for population studies of geographically diverse groups (Batzer et al. 1991, 1993, 1994, in press, Novick et al. 1992, in press). In light of all this evidence, even if full exponential amplification is halted by limiting factors, the existence of a small number of subfamilies provides strong support for the master gene model.

An interesting related issue concerns the potential factors that regulate *Alu* copy number. Because *Alu* insertions are for the most part deleterious, it is logical to expect a maximal number of insertions after which the genetic burden or load would not tolerate additional *Alu* insertions. The 100-fold decrease in *Alu* insertion rate over the last 40–50 million years may be indicative of a limit to the number of *Alu* copies that our gene pool can withstand.



**Figure 4.** Agarose gel electrophoresis and codominant segregation of polymorphic *Alu* PCR products for the angiotensin converting enzyme locus in a family. The pedigree studied is represented on top of the figure, and the electrophoretic separation is found immediately under each individual on the pedigree. Lanes 1 and 9 represent molecular size standard *Hae* III digested  $\phi$ X174 DNA. Arrow A indicates the insertion band (490 bp for this locus), arrow B points to the lack of insertion band (190 bp for this locus), and arrows C and D represent primer-dimer and primer, respectively.

The force responsible for establishing and maintaining such a limit may be a selection against individuals experiencing *de novo* insertions, especially if they significantly disrupt the function of critical genes. In theory, once the number of *Alu* copies in the human gene pool reached this limit, only a few *de novo* insertions could become fixed in the population, replacing the *Alu* copies lost, for example, during unequal crossing over.

### ***Alu* insertions and the genomic environment**

Do *Alu* elements have any function? How does their presence affect neighboring sequences and the genome as a whole? Transposable ele-

ments were initially considered mere parasites of the genome with no major effect on the organism (Orgel and Crick 1980). This perception resulted in what is known as the selfish DNA hypothesis (Orgel and Crick 1980), wherein repeated DNA sequences, and transposable elements in particular, have no cellular function or genetic effect but to increase their copy number. The phrase *selfish DNA* has become so popular that it is included in *The Oxford English Dictionary* (1989), which defines it as "genetic material tending to be perpetuated or to spread although of no effect on the phenotype."

Initially, transposable elements were viewed by some as being either molecular fossils or inconsequential

both to the cell and to the evolution of the species. New molecular evidence regarding the structure and evolution of this class of sequences, however, suggests that although repeated DNA sequences may have originated as innocuous elements, they can have dramatic effects on genome organization, function, and evolution (Makalowski et al. 1994). Although the notion of *Alu* elements as mutagens causing deleterious effects is well recognized, there are also innumerable examples of *Alu* insertions involved in normal DNA structure and function. To cite a few examples, representative and recent studies have demonstrated the role of *Alu* elements as DNA organizers (Englander and Howard 1995), enhancers (Almenoff et al. 1994), and silencers (Hanke et al. 1995) in gene expression. Over the past decade, the list of studies pointing to roles performed by SINES is extensive (see von Sternberg et al. 1992 for review of the literature). Therefore, whether or not *Alu* elements and SINES in general, are considered selfish DNA, they can result in functional sequences.

As an alternative to the selfish-DNA hypothesis, we proposed a model called genome canalization (von Sternberg et al. 1992). Genome canalization is the process whereby once a pattern of genomic organization is established, all further evolutionary changes must adapt to the prevailing genetic context. The basic concept of genome canalization is not a novel one, coevolution being the underlining principle. In the same way that birds and plants coevolved specific peak and flower morphology for their mutual benefit, so it is that different types of DNA coevolved to perform a function.

At the DNA level, for example, there is an intimate relationship between genes and their transcription regulatory sequences brought about by the process of molecular coevolution. *Alu* elements and other SINES are just one source of genetic material that is often placed in physical proximity to other DNA sequences, dramatically fueling the coevolutionary process.

The various types of mutations (transposition events being among

the most dramatic) would be channeled by the interaction of the chromosomal segments at all levels. In this model, transposable elements cannot be said to have evolved to meet some specific adaptive need, but they have acquired roles fortuitously due to their widespread genomic distribution and novel structural properties (Brosius 1991). As a result, transposable elements can now be regarded as having potential functions, although not adaptive ones (Doolittle 1989).

Doolittle (1989) has defined the term *exaptive function* as "features that now enhance fitness but were not built by natural selection for their current roles...." From this perspective, mobile sequences could well have originated in a selfish manner; however, once distributed chromosomally, individual elements diverged and their genetic effects resulted in particular functions. Once such functions arose, they were then capable of being under selective pressure (Brosius 1991). In other words, with time and the action of selective evolutionary forces, the exaptive function may have evolved into an adaptive one.

Table 1 illustrates the retroposon mode of insertion, potential genomic insertional sites, immediate consequences on gene function and organization, and some of the possible adaptive outcomes within structural genes and regulatory sequences. *Alu* insertions are capable of affecting the structure and function of the rest of the genome in many ways.

This finding is not surprising considering that these elements contain genetic information, including transcription regulatory sequences, that insert themselves indiscriminately all over the genome. Subsequent to insertion, coevolution with different types of neighboring sequences ultimately may yield diverse novel functions for the DNA region in question.

Table 1 focuses on the immediate effects on gene structure and function of *Alu* insertions in different genomic environments. Other factors like genetic drift, migration, or polymorphisms near the insertional sites act at the population level and can, in turn, influence the fate of *Alu* elements as well.

**Table 1.** Potential effects of *Alu* insertions.

Insertion	Potential effect on gene organization and function
Into promoters	Trans-acting factors and/or RNA polymerase cannot recognize the promoter sequences leading to transcription failure.
Between promoters	The distance between promoters is modified and their function altered or obliterated preventing transcription initiation.
Into enhancers	The action of enhancers over promoters is altered affecting transcription levels.
Into exons	A new product is produced with similar or different function(s).  Rearrangements are promoted leading to addition/deletion and frame shift mutations.  A cryptic splice site is activated creating a new gene product.  The <i>Alu</i> element is removed.
Into introns	The splice site(s) are modified generating a new gene product with similar or different function(s).  The <i>Alu</i> element is spliced out.
Into heterochromatic domains	New <i>Alu</i> elements tend to cluster in the area affecting expression of nearby genes.

## ***Alu* insertions and human biology**

Some representative examples of *Alu* insertions causing human disorders are:

- **Hemophilia**—a series of disorders that are caused by mutations in any of a number of different blood-clotting factors. A de novo insertion of an HS *Alu* family member into the blood coagulation factor IX locus has resulted in hemophilia (Vidaud et al. 1993). This *Alu* repeat inserted into exon V of the factor IX gene results in the generation of a premature stop codon. This insertion gives rise to a nonfunctional factor IX, which is inherited as an X-linked disorder.

- **Meningiomas**—tumors of the membranes that surround the central nervous system. Meningiomas are frequently associated with complete or partial deletions of chromosome 22, and the gene most frequently affected is the *c-sis* gene, which encodes one of the two polypeptide subunits of the platelet-derived growth factor (Dalla Favera et al. 1982). The deletion of an *Alu* element located in the fifth intron of *c-sis* is associated with the development of this type of tumor (Smidt et

al. 1990). Therefore, this human malady represents an *Alu* element necessary for normal gene function.

- **Hereditary angioneurotic edema**—a genetic disorder caused by the deficiency of an inhibitor called C1 (C1 INH). Seventeen *Alu* repeat sequences were identified within the C1 INH gene (Carter and Fothergill 1989), and one study demonstrated that partial deletions, observed in two families with hereditary angioneurotic edema, are the result of *Alu-Alu* recombination by unequal crossing-over within the C1 INH gene (Ariga et al. 1990). The most frequent breakpoint for recombination lies on the left half of the *Alu* elements between two A and B boxes of the RNA polymerase III promoter.
- **Familial hypercholesterolemia**—an autosomal dominant disease caused by mutations in the gene for the low density lipoprotein (LDL) receptor. A defective LDL receptor in a member of an affected family was found to be 50,000 daltons larger than normal. The anomalous protein was found to be the result of a 14-kilobase duplication due to unequal crossing-over between homologous *Alu* elements in introns 1 and 8 (Lehrman et al. 1987). Again the most frequent breakpoints were found in the area of the RNA poly-

merase III promoter.

- Gyrate atrophy of the choroid and retina—an autosomal recessive chorioretinal degeneration caused by deficiency of ornithine k-aminotransferase. In one patient with typical symptoms and consanguineous parents, the molecular defect was found to be a 142-nucleotide insertion in the mature ornithine k-aminotransferase mRNA due to a splice-mediated insertion mutation of an *Alu* element in opposite orientation to the affected gene. This *Alu* element is normally located in intron 3 of the gene (Mitchell et al. 1991).

- Severe combined immune deficiency—a disorder caused by a dysfunction of the gene that codes for adenosine deaminase (ADA). Two independent cases of unrelated origin showed that the defect was due to a 3.2-kilobase deletion in the promoter and first exon of ADA. This deletion was confirmed to be the result of homologous recombination between *Alu* repeats. The breakpoint was mapped to the left arm of the *Alu*, in the area of the polymerase III promoter (Berkvens et al. 1990).

- Neurofibromatosis type 1 (NF1)—a common autosomal dominant disorder characterized by neurofibromas, other cutaneous disorders, bone deformities, and learning disabilities. A mutation found in an affected adult male is a de novo *Alu* insertion into an intron of the NF1 gene (Wallace 1991). The target site for insertion was found to be a 26-base pair A+T rich region that is known to be a preferential site for *Alu* insertions. The presence of the element causes intron sliding in which the following exon is removed along with the intron during splicing and the reading frame is shifted (Wallace 1991).

- Glanzmann thrombasthenia (GT)—an autosomal recessive bleeding disorder due to a defect in the platelet fibrinogen receptor, which causes an inability of the platelets to either bind to fibrinogen and aggregate. Both binding to fibrinogen and aggregation are crucial for normal coagulation. In one affected family, the molecular defect was identified as a 1-kilobase deletion immediately preceding a 15-kilobase inversion

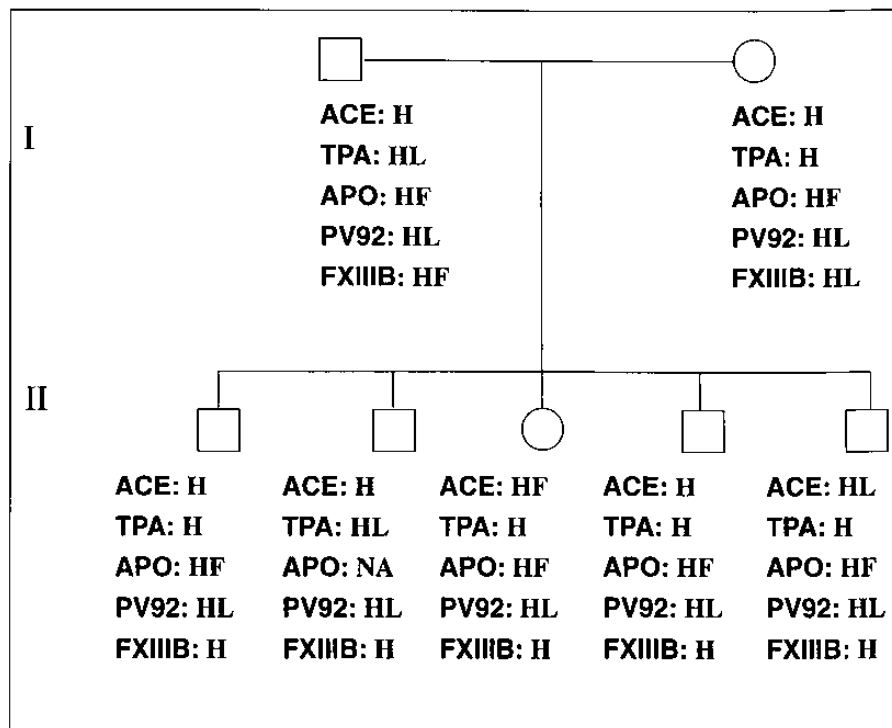


Figure 5. Summary of scores obtained from the members of the pedigree represented in Figure 4 after the *Alu* PCR polymorphism assay for each of the five loci studied. HF: homozygous for the insertion; H: heterozygous; HL: homozygous for the lack of insertion.

due to breaks in two antisense (opposite orientations) and one sense *Alu* sequences within the gene of one of the glycoproteins (GPIIb) that constitute the receptor (Lu and Bray 1993).

Although *Alu* insertions or deletions in specific genes have been found to result in the described diseases, there are other types of mutations (e.g., base substitutions) that can cause many of the above-mentioned disorders as well.

In addition, *Alu* elements are implicated in the evolution of new genes. An example can be found in glycophorins A (GPA) and B (GPB)—membrane sialoglycoproteins of human erythrocytes. When their amino acid and cDNA sequences are compared, there is one area where GPA and GPB are identical and another where they are entirely different (Kudo and Fukada 1989). The transition site from homologous to nonhomologous sequences occurs within an *Alu* repeat. Kudo and Fukada (1989) postulate that GPB originated by duplication of GPA and homologous recombination between two adjacent *Alu* elements,

which may have also been involved in the original duplication event.

### *Alu* elements as research tools

*Alu* repeats as markers for human population genetics and DNA fingerprinting. We have already described the way in which subfamilies of *Alu* elements might have arisen from sequential mutations of a master gene. Recently inserted *Alu* elements (like some HS-1, HS-2, and Sb-2 subfamily members) are only found in some of the insertion sites in a given population (Batzer et al. 1995). Therefore, they represent insertion polymorphisms within populations. Given that *Alu* elements are generally highly stable after integration (Sawada and Schmid 1986), and each element is absent from nonhuman primates, the lack of insertion can be considered the ancestral state and the presence of the *Alu* element in that location the recent character state.

We have developed a polymerase chain reaction (PCR) assay to detect the presence or absence of *Alu* insertions (Batzer et al. 1991). PCR reproduces in vitro the events of



DNA replication in a cyclic manner, resulting in the exponential amplification of the original target sequence (Saiki et al. 1988). The principle of the *Alu* polymorphism assay is summarized in Figure 3. The assay uses oligonucleotide primers directed to the 5' and 3' single-copy sequences flanking the specific *Alu* element to be studied. This assay generates three potential outcomes when the PCR products are separated by electrophoresis on an agarose or acrylamide gel. Homozygous individuals for the *Alu* insertion amplify, for example, a large fragment that represents the *Alu* insertion between the two primers; an individual homozygous for the lack of insertion shows only a small fragment; and a heterozygous individual amplifies both products—the small fragment from the allele that lacks the insertion and the large fragment from the insertion allele (see lower panels in Figure 3).

When this assay is used to screen geographically distinct populations, a genetic profile of each group can be constructed and compared. All the characteristics of *Alu* elements described thus far (mechanism of retroposition, relatively nonspecific site of insertion, rate of amplification, and variability of the oligo dA track) suggest that it is unlikely two populations of unrelated origin have the same insertion frequencies for several *Alu* insertions. The stability and dimorphic distribution of *Alu* elements make them ideal markers for population studies. For example, we recently presented evidence for an African origin of modern human populations using only four polymorphic *Alu* insertion loci (Batzer et al. 1994), and we demonstrated admixture involving African Colombian and American Indian populations (Novick et al. in press) as well as African Caribbean and African American groups (Batzer et al. in press).

In the area of DNA fingerprinting, multiple approaches are currently being used and new strategies are in constant development. Restriction fragment length polymorphisms or RFLPs (Kan and Dozy 1978) and hypervariable minisatellite regions (Jeffreys et al. 1985) are the most widely used DNA typing methods. Recently, polymorphic

*Alu* insertions have been employed successfully in human identification studies (Novick et al. 1994).

Many technologies, aimed at human identification, have been developed. These technologies include the use of monoclonal antibodies for the identification of body fluids (Martin and Parkin 1988), protein polymorphisms (Hobart 1979), human leukocyte antigen (HLA) typing (Ishitani and Hirota 1988), short tandem repeat PCR polymorphisms (Edwards 1991), amplified fragment length polymorphisms (Eisenberg and Maha 1991), and minisatellite variant repeat-PCR (Jeffreys et al. 1991).

These techniques can be used to answer different types of questions, and the degree of variability observed by the different procedures is likely to limit their optimal range of usage. For example, ABO typing may not be a conclusive method to differentiate between individuals, because different individuals may share the same blood group, while variable number tandem repeats, on the other hand, would not be ideal to study population differences, because the same allelic variant can arise multiple times in different populations and their hypervariability generates so many alleles that no clear differences between populations may be evident.

The nature of *Alu* insertions makes it improbable that the same insertion occurred independently within human evolutionary time at exactly the same genomic location in two different individuals. In addition, there is no known mechanism to precisely delete just the repetitive elements without deleting flanking sequences or leaving residual *Alu* sequences behind. Therefore, polymorphic *Alu* insertions represent a distinct, easy-to-measure genetic change resulting from an event that occurred one time within the human lineage. These characteristics make the polymorphic *Alu* insertions valuable tools for human population genetic studies and DNA typing, while providing information for the window of genetic variability left by highly variable sequences on the one hand and more conserved single and multiple copy genes on the other.

**The use of polymorphic *Alu* insertions in paternity testing.** We applied the principle of PCR-based *Alu* polymorphism testing to ascertain paternity using five polymorphic *Alu* insertions. The five HS *Alu* elements examined are unique to the human lineage. Their genomic localizations have been described (reviewed in Novick et al. 1994). In one study, we compared our strategy with the two approaches most widely used in paternity testing: HLA molecules and RFLP analysis.

The HLA complex is the most polymorphic set of genetic markers in humans, so it is a powerful system to resolve cases of disputed paternity (Lamm et al. 1983). RFLP analysis is based on differential digestion of the DNA with a restriction enzyme. This analysis results in a distinctive banding pattern that reveals not only the virtually unique genetic profile of an individual but also allows the parental origin of each band to be determined (Helminen et al. 1988).

Figure 4 displays the outcome of the analysis of one family for one locus. The results of the paternity testing using HS *Alu* polymorphism for the five loci in the same family is shown in Figure 5. To determine the likelihood of parentage of the two parents with respect to the five offspring of the pedigree, the paternity index and paternity probability were calculated using standard formulas as previously described (Novick et al. 1994).

The RFLP and HLA analysis of this pedigree yielded a probability of paternity of 99.99% and 85%, respectively. According to the HS *Alu* polymorphism assay, the probability of paternity obtained was 89%, and the probability of maternity was 96%. There are hundreds of potential polymorphic *Alu* markers for this type of analysis. Currently, our laboratories are in the process of characterizing additional polymorphic *Alu* insertions. Five to ten additional polymorphic *Alu* markers would allow probability determinations in the order of 99.99%.

## Conclusions

DNA structure was initially viewed as a chain of rigid building blocks

and the sequence DNA-to-RNA-to-protein was the sole and central dogma describing gene function. Unexpected discoveries in the field of gene structure and function have opened new perspectives, and old paradigms are replaced by continuously emerging hypotheses. Within this context, transposable elements, and *Alu* sequences in particular, constitute a large portion of the genome. Their insertion is not entirely random and, rather than being useless sequences, they represent a dramatic source of genetic variation.

Like any other type of mutation, *Alu* sequences are in most cases deleterious but not useless to the population as a whole. As retrotransposons and facilitators of unequal crossing over, they reshuffle and duplicate the DNA, thereby generating a significant amount of genetic variability while providing genetic flexibility and increasing the population's survival in a changing environment. As a result, there are innumerable cases in which *Alu* elements and SINEs, in general, function in normal gene structure and activity. Transposable elements constitute a remarkably active component of a fluid and constantly changing gene pool, having only partially understood functions, and can be reckoned as a force helping to shape the evolutionary path of the entire genome.

## Acknowledgments

This research was supported by grants from the National Institutes of Health (RR08205) to Rene J. Herrera and (R01 HG 00770) to Prescott L. Deininger. We would like to thank Ralph L. Diaz-Wagner and Frank Verde for their editorial assistance.

## References cited

- Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. 1989. Molecular biology of the cell. New York: Garland Publishing.
- Almenoff JS, Jurka J, Schoolnik GK. 1994. Induction of heat-stable enterotoxin receptor activity by a human *Alu* repeat. *Journal of Biological Chemistry* 269: 16610-16617.
- Ariga T, Carter PE, Davis AE III. 1990. Recombinations between *Alu* repeat sequences that result in partial deletions within the C1 inhibitor gene. *Genomics* 8: 607-613.
- Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselts TH, Hoppens CL, Deininger PL. 1990. Structure and variability of recently inserted *Alu* family members. *Nucleic Acids Research* 18: 6793-6998.
- Batzer MA, Gudi VA, Mena JC, Foltz DW, Herrera RJ, Deininger PL. 1991. Amplification dynamics of human-specific (HS) *Alu* family members. *Nucleic Acids Research* 19: 3619-3623.
- Batzer MA, et al. 1993. *Alu* repeats as markers for human population genetics. Pages 49-57 in *Proceedings from the Fourth International Symposium of Human Identification*. Madison (WI): Promega Corp. Publishing.
- Batzer MA, et al. 1994. African origin of human-specific polymorphic *Alu* insertions. *Proceedings of the National Academy of Sciences of the United States of America* 91: 12288-12292.
- Batzer MA, et al. 1995. Dispersion and insertion polymorphism in two small subfamilies of recently amplified *Alu* repeats. *Journal of Molecular Biology* 247: 418-427.
- Batzer MA, et al. In press. Genetic variation of recent *Alu* insertions in human populations. *Journal of Molecular Evolution*.
- Berkvens TM, van Ormondt H, Gerritsen FJA, Meera Khan P, van der Eb AJ. 1990. Identical 3250-bp deletion between two *Alu* I repeats in the ADA genes of unrelated ADA-SCID patients. *Genomics* 7: 486-490.
- Britten RJ. 1994. Evidence that most human *Alu* sequences were inserted in a process that ceased about 30 million years ago. *Proceedings of the National Academy of Sciences of the United States of America* 91: 6148-6150.
- Britten RJ, Kohne DE. 1968. Repeated sequences in DNA. *Science* 161: 529-540.
- Britten RJ, Stout DB, Davidson EH. 1989. The current source of human *Alu* retrotransposons is a conserved gene shared with old world monkey. *Proceedings of the National Academy of Sciences of the United States of America* 86: 3718-22.
- Brosius J. 1991. Retrotransposons—seeds of evolution. *Science* 251: 753.
- Carter PE, Fothergill JE. 1989. The complete nucleotide sequence of the gene for human C1 INH. *Complement Inflammation* 6: 37.
- Dalla Favera R, Gallo RC, Giallongo A, Croce CM. 1982. Chromosomal localization of the human homolog (*c-sis*) of the simian sarcoma virus oncogene. *Science* 218: 686-688.
- Daniels GR, Fox M, Lowenstein D, Schmid C, Deininger PL. 1983. Species specific homogeneity of the primate *Alu* family of repeated DNA sequence. *Nucleic Acids Research* 11: 7579-7593.
- Deininger PL. 1989. SINEs: short interspersed repeated DNA elements in higher eucaryotes. Pages 619-636 in Howe M, Berg D, eds. *Mobile DNA*. Washington (DC): ASM Press.
- Deininger PL, Batzer MA. 1993. Evolution of retrotransposons. *Evolutionary Biology* 27: 157-196.
- Deininger PL, Batzer MA, Hutchinson C III, Edgell MH. 1992. Master genes in mammalian repetitive DNA amplification. *Trends in Genetics* 8: 307-312.
- Doolittle WF. 1989. Hierarchical approaches to genome evolution. *Canadian Journal of Philosophy Supplement* 14: 101-133.
- Economou EP, Bergen AW, Warren AL, Antonarakis SE. 1990. The polydeoxyadenylate tract of *Alu* repetitive elements is polymorphic in the human genome. *Proceedings of the National Academy of Sciences of the United States of America* 87: 2951-2954.
- Edwards A. 1991. DNA typing with trimeric and tetrameric tandem repeats: polymorphic loci, detection systems and population genetics. Pages 31-52 in *Proceedings from the Second International Symposium of Human Identification*. Madison (WI): Promega Corp. Publishing.
- Eisenberg M, Maha G. 1991. AMPFLP analysis in parentage testing. Pages 129-154 in *Proceedings from the Second International Symposium of Human Identification*. Madison (WI): Promega Corp. Publishing.
- Englander EW, Howard BH. 1995. Nucleosome positioning by human *Alu* elements in chromatin. *Journal of Biological Chemistry* 270: 10091-10096.
- Fuhrman SA, Deininger PL, LaPorte P, Friedmann T, Geiduschek EP. 1981. Analysis of transcription of the human *Alu* family ubiquitous repeating element by eukaryotic RNA polymerase III. *Nucleic Acids Research* 9: 6439-6456.
- Hanke JH, Hambor JE, Kavathas P. 1995. Repetitive *Alu* elements form a cruciform structure that regulates the function of the human CD8 alpha T cell-specific enhancer. *Journal of Molecular Biology* 246: 63-75.
- Helminen P, Ehnholm C, Lokki M, Jeffreys A, Peltonen L. 1988. Application of DNA fingerprints to paternity determination. *Lancet* 1: 574-576.
- Hobart MJ. 1979. Genetic polymorphism of human plasminogen. *Annals of Human Genetics* 42: 419-423.
- Houck CM, Rinehart FP, Schmid CW. 1979. A ubiquitous family of repeated DNA sequences in the human genome. *Journal of Molecular Biology* 132: 289-306.
- Ishitani A, Hirota T. 1988. Personal identification by HLA typing of cultured fibroblast derived from cadaveric tissues. Pages 59-63 in Mayr WR, ed. *Advances in forensic haemogenetics*. Vol. 2. Berlin (Germany): Springer-Verlag.
- Jagadeeswaran P, Forget BG, Weissman SM. 1981. Short interspersed repetitive DNA elements in eucaryotes: transposable DNA elements generated by reverse transcription of RNA pol III transcripts? *Cell* 26: 141-142.
- Jeffreys AJ, Wilson V, Thein SL. 1985. Hypervariable "minisatellite" regions in human DNA. *Nature* 314: 67-73.
- Jeffreys AJ, MacLeod A, Tamaki K, Neil D, Monckton DG. 1991. Minisatellite repeat coding as a digital approach to DNA typing. *Nature* 354: 204-209.
- Kan YW, Dozy A. 1978. Polymorphism of DNA sequences adjacent to human B-globin structural gene: relationship to the sickle mutation. *Proceedings of the National Academy of Sciences of the United States of America* 75: 5631-5635.
- Korenberg JR, Rykowski MC. 1988. Human genome organization: *Alu*, LINE and the molecular structure of metaphase chromosome bands. *Cell* 53: 391-400.
- Kudo S, Fukuda M. 1989. Structural organization of glycophorin A and B genes: glycophorin B gene evolved by homologous recombination at *Alu* repeat sequences. *Proceedings of the National Academy of*

- Sciences of the United States of America 86: 4619-4623.
- Lamm LU, Gurtler H, Hansen HE. 1983. The HLA system. Pages 381-396 in Inclusion probabilities in parentage testing. Arlington (VA): American Association of Blood Banks.
- Lehrman MA, Goldstein JL, Russell DW, Brown MS. 1987. Duplication of seven exons in LDL receptor gene caused by *Alu-Alu* recombination in a subject with familial hypercholesterolemia. Cell 48: 827-835.
- Lu L, Bray PF. 1993. Homologous recombination among three intragene *Alu* sequences causes an inversion-deletion resulting in the hereditary bleeding disorder glanzmann thrombasthenia. American Journal of Human Genetics 53: 140-149.
- Makalowski W, Mitchell GA, Labuda D. 1994. *Alu* sequences in the coding regions of mRNA: a source of protein variability. Trends in Genetics 10: 188-193.
- Martin PD, Parkin BH. 1988. The use of monoclonal antibodies in forensic sciences. Pages 284-297 in Mayr WR, ed. Advances in forensic haemogenetics. Vol. 2. Berlin (Germany): Springer-Verlag.
- Martin MA, Bryan T, Rasheed S, Khan AS. 1981. Identification and cloning of endogenous retroviral sequences present in human DNA. Proceedings of the National Academy of Sciences of the United States of America 78: 4892-4896.
- Mathias SL, Scott AF, Kazazian HH, Boeke JD, Gabriel A. 1991. Reverse transcriptase encoded by a human transposable element. Science 254:1808-1810.
- Mitchell GA, Labuda D, Fontaine G, Saudubray JM, Bonnefont JP, Lyonnet S, Brody LC, Steel G, Obie C, Valle D. 1991. Splice-mediated insertion of an *Alu* sequence inactivates ornithine  $\delta$ -aminotransferase: a role for *Alu* elements in human mutation. Proceedings of the National Academy of Sciences of the United States of America 88: 815-819.
- Mount SM, Rubin GM. 1985. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. Molecular and Cellular Biology 5: 1630-1638.
- Novick GE, Gonzalez T, Garrison J, Novick CC, Batzer MA, Deininger PL, Herrera RJ. 1992. The use of polymorphic *Alu* insertions in human DNA fingerprinting. Pages 284-291 in Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ, eds. DNA fingerprinting: state of the science. Basel (Switzerland): Verlag.
- Novick GE, Menendez CM, Novick CC, Duncan G, Yunis J, Yunis E, Deininger PL, Batzer MA, Herrera RJ. 1994. The use of polymorphic *Alu* insertions as a new methodological alternative in human paternity testing and child identification. International Pediatrics 9: 60-68.
- Novick GE, et al. In press. Polymorphic human specific *Alu* insertions as markers for human identification. Electrophoresis.
- Orgel LE, Crick FHC. 1980. Selfish DNA: the ultimate parasite. Nature 284: 604-607.
- The Oxford English Dictionary. 1989. Page 924 in The Oxford English Dictionary. 2nd ed. Vol. XIV. Oxford (UK): Oxford University Press.
- Paulson KE, Deka N, Schmid C, Misra R, Schindler C, Rush M, Kadyk L, Leinwand L. 1985. A transposon-like element in human DNA. Nature 316: 359-363.
- Rinehart FP, Ritch TG, Deininger PL, Schmid CW. 1981. Renaturation rate studies of a single family of interspersed repeated sequences in human deoxyribonucleic acid. Biochemistry 20: 3003-3010.
- Rogers J. 1985. The origin and evolution of retrotransposons. International Review of Cytology 93: 187-279.
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-491.
- Sawada I, Schmid CW. 1986. Primate evolution of the  $\alpha$ -globin gene cluster and its *Alu*-like repeats. Journal of Molecular Biology 192: 693.
- Schmid CW, Maraia R. 1993. Transcriptional regulation and transpositional selection of active SINE sequences. Current Opinions in Genetics Development 2: 874-882.
- Shen MR, Batzer MA, Deininger PL. 1991. Evolution of the master *Alu* gene(s). Journal of Molecular Evolution 33: 311-322.
- Smidt MJ, Kirsch I, Ratner L. 1990. Deletion of *Alu* sequences in the fifth *c-sis* intron in individuals with meningiomas. Journal of Clinical Investigation 86: 1151-1157.
- Ullu E, Tschudi C. 1984. *Alu* sequences are processed 7SL RNA genes. Nature 312: 171-172.
- Ullu E, Weiner AM. 1985. Upstream sequences modulate the internal promoter of the human 7SL RNA gene. Nature 318: 371-374.
- Vidaud D, Vidaud M, Bahnak BR, Siguret V, Sanchez SG, Laurin Y, Meyer D, Goossens M, Lavergne JM. 1993. Hemophilia B due to a *de novo* insertion of a human-specific *Alu* subfamily member within the coding region of the factor IX gene. European Journal of Human Genetics 1: 30-36.
- von Sternberg RM, Novick GE, Gao GP, Herrera RJ. 1992. Genome canalization: the coevolution of transposable and interspersed repetitive elements with single copy DNA. Genetica 86: 215-246.
- Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS. 1991. A *de novo* *Alu* insertion results in neurofibromatosis type 1. Nature 353: 864-866.
- Walter P, Blobel G. 1982. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature 299: 691-698.

# Dial -a- tree.



Ok, you watch the news. You're worried about the future. Who is going to protect the environment for you and your kids?

But what can you do? You're pretty busy these days. You don't have time to go out into the woods and play forest ranger.

Relax. AMERICAN FORESTS has the answer. Sponsor the planting of trees in a Global ReLeaf Forest.

Pick up your phone right now and call 1-800-873-5323. Each tree costs just \$1. So go ahead and dial-a-tree.

Technically, you punch in the numbers. But punch-a-tree doesn't sound as good.

AMERICAN FORESTS



1-800-873-5323