*Proc. Natl. Acad. Sci. USA*
Vol. 86, pp. 3718–3722, May 1989
Genetics

# The current source of human *Alu* retroposons is a conserved gene shared with Old World monkey

(primate/repeated sequences/conservation/drift/DNA)

Roy J. Britten*†‡§, David B. Stout*, and Eric H. Davidson‡

*Kerckhoff Marine Laboratory, California Institute of Technology, 101 Dahlia Avenue, Corona del Mar, CA 92625; ‡Division of Biology, California Institute of Technology, Pasadena, CA 91125; and †Carnegie Institution of Washington, Washington, DC 20005

*Contributed by Roy J. Britten, February 6, 1989*

ABSTRACT **A significant fraction of human *Alu* repeated sequences are members of the precise, recently inserted class. A cloned member of this class has been used as a probe for interspecies hybridization and thermal stability determination. The probe was reassociated with human, mandrill, and spider monkey DNA under conditions such that only almost perfectly matching duplexes could form. Equally precise hybrids were formed with human and mandrill DNA (Old World monkey) but not with spider monkey DNA (New World). These measurements as well as reassociation kinetics show the presence in mandrill DNA of many precise class *Alu* sequences that are very similar or identical in quantity and sequence to those in human DNA. Human and mandrill are moderately distant species with a single-copy DNA divergence of about 6%. Nevertheless, their recently inserted *Alu* sequences arise by retroposition of transcripts of source genes with nearly identical sequences. Apparently a gene present in our common ancestor at the time of branching was inherited and highly conserved in sequence in both the lineage of Old World monkeys and the lineage of apes and man.**

Since its discovery about 25 years ago (1, 2) little has been learned of the origin of highly repetitive DNA, but a beginning is now being made for the human *Alu* repeat. Ultimately we may learn why nearly a million *Alu* repeats have been interspersed throughout the DNA all during the course of primate evolution. The *Alu* repeat of primate DNA is short [281 nucleotides plus a terminal poly(A) stretch], interspersed about every 4 kilobases (kb) throughout the human genome (3), and has been inserted in the DNA by retroposition (4). It is now possible to describe the underlying process as the insertion of DNA copies of RNA transcripts of "source" genes (5).

The evidence that many *Alu* sequences are copies of conserved source gene sequences depends primarily on the classification of *Alu* repeats by their sequence relationship. Initially the *Alu* repeated sequences were divided on the basis of their degree of divergence from each other (6) into three subsets ("conserved, majority, and divergent") and a new consensus was derived for the sets that are closely similar to each other (6, 7). Since it is apparently not the set of *Alu* sequences that is conserved but their source gene we use the term "precise" rather than "conserved" consensus. The precise consensus sequence is probably identical to the sequence of the source gene that is responsible for recent *Alu* insertions (5).

By sequence alignment and examination of shared substitutions at a set of diagnostic positions, an improved subdivision of *Alu* sequences into classes may be made (5, 8, 9). Although five classes clearly exist, recent studies of >300 *Alu* sequences (unpublished) suggest that six classes or more may ultimately be established. The overlapping pattern of the substitutions shared between classes indicates that there has been an evolutionary series of source genes, each dominating the insertion of *Alu* repeats for a time and differing from the previous source by a set of mutations. Before the mutations, in almost every case, each of the nucleotides in the set matched those of 7SL RNA (5, 8). After each group of mutations most of the resulting new nucleotides were preserved right up to the present. In other words, though there are periods when mutations occur at a higher rate, the sources appear to be conserved in sequence during most of the evolutionary time course (5).

In contrast to the sources, *Alu* repeats once inserted in the genome are not conserved, and most positions in their sequences drift at about the same rate as single-copy DNA (10–12) or about 0.15% per million years (13). This contrast is particularly clear for the 25 CpG dinucleotides that are conserved in the source genes. After insertion of *Alu* sequences, mutations leading to the loss of the CpG dinucleotides occur at 10 times the rate of mutations at other positions (5), apparently as a result of instability due to methylation (14).

The measurements described here show that mandrill (Old World monkey) DNA includes many *Alu* sequences that are very similar or identical to many *Alu* sequences in human DNA. Apparently a gene acting as a source of *Alu* sequence retroposition carried by our common ancestor was inherited and highly conserved in both the lineages leading to Old World monkeys and to apes and man.

## METHODS

A clone in bacteriophage M13 of 780 nucleotides of the $\delta\beta$-globin intergenic region of gorilla DNA containing a recently inserted *Alu* repeat (15) was a gift from G. Trabuchet (University of Lyon, France). It differs in six positions from the precise consensus (6, 7). To use as a probe it was $^{32}$P labeled by extension with *Pol* I from a sequencing primer and cut with *Kpn* I. After electrophoresis on an alkaline gel, the purified fragment was self-incubated overnight in 0.48 M neutral sodium phosphate buffer (PB) at 55°C and passed over hydroxyapatite in 0.12 M PB at 50°C to remove any duplexes. The unbound fraction was used as the labeled probe in hybridizations with 10 $\mu$g of sonicated driver DNA from the three species in 20 $\mu$l for 1 hr. Controls showed no self duplex formation. Hybridization with human DNA reached only 65–75%, presumably due to contamination with short fragments and non-*Alu* regions of the M13 clone. Incubation was at 55°C or 60°C in 2.0 M tetraethylammonium chloride (TEACl)/0.013 M PB, conditions in which long

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: $t_m$, melting temperature.

§To whom reprint requests should be addressed at: Kerckhoff Marine Laboratory, California Institute of Technology, 101 Dahlia Ave., Corona del Mar, CA 92625.

native DNA melts at 69°C independent of base composition. Binding to hydroxyapatite was done in this buffer at 50°C; this was followed by wash in 0.013 M PB to remove the TEACl, wash with 0.12 M PB at 50°C, and elution with 0.12 M PB as the temperature was raised. The native duplex of the probe (replicative form) is denatured and elutes from hydroxyapatite under these conditions at 93°C (5).

## RESULTS

**The Probe for the Precise Class.** The precise class [identified as class IV (5), b (8), or A (9)] has a small average divergence of about 3% from its consensus as shown in Fig. 1. It is likely that its members have been inserted over the last 20 million years, recent in comparison with the other classes. Two examples are known that may be interpreted as human DNA polymorphism for the presence or absence of an *Alu* sequence in a homologous location (16, 17). In one case an *Alu* was found in a clone from the genome of a lymphoma cell line that is absent from 59 other human genomes examined (16). In a second case an *Alu* was present in one insert in the Maniatis human DNA library (18) but absent from another insert from the same library containing the homologous region (17). It has been pointed out that these two sequences share five substitutions that differ from the precise consensus (19) and may have been produced by a new human source gene. No significant combination of any of these five substitutions has been found in a search of >300 *Alu* sequences from GenBank, so no other members of this potential family have been identified. An example of a transposition of an older (class II) sequence has been observed after severe ultraviolet exposure of tissue culture cells (20). All three other cases of recent insertion belong to the precise class, including the one described below.

An *Alu* sequence is present in gorilla DNA (15) that is absent from the homologous location in the DNA of human or chimpanzee and may not be present in all individual gorillas. This *Alu* sequence, inserted into the DNA of some gorilla lineages after the human and gorilla lineages separated (15), differs in only 6 nucleotides from the precise consensus (5–7). An M13 clone containing this sequence is a useful probe for the *Alu* sequences of the precise class. We previously showed that when this probe is hybridized with total human DNA, about 25% forms very precise duplexes (5), indicating that many *Alu* repeats have been recently inserted.

**Measurement of the Precise Class *Alu* Sequences by Hybridization.** To examine specifically the precise set, hybridization may be done at "high criterion" by incubation at a temperature just below the melting point of precise hybrids. Under these conditions the probe will hybridize only with nearly perfectly matching *Alu* sequences and fails to hybridize if they are absent. An example is shown in Fig. 2 where the incubation was done 9°C (Fig. 2*A*) and 14°C (Fig. 2*B*) below the melting temperature ($t_m$) of long perfect duplexes. The results of these measurements show that mandrill (*Mandrillus sphinx*) DNA contains a set of copies of the *Alu* repeat that are as close in sequence to the precise consensus as are those in human DNA.

In Fig. 2*A*, the duplex $t_m$ for both is about 90°C as measured by elution from hydroxyapatite, whereas controls (5) show that the native 780-nucleotide-long duplex of this cloned probe melts at 93°C. The $t_m$ of DNA duplex is reduced by short fragment length (L) by 500/L (21). The *Alu* sequence is about 300 nucleotides long and thus the hybrids would be expected to melt about 1°C lower because of $t_m$ reduction due to shorter length. In addition, the probe differs in 6 nucleotide positions from the precise consensus giving another 2°C reduction in $t_m$ (1% divergence reduces the $t_m$ about 1°C). Thus the $t_m$ of hybrids with the precise consensus sequence is expected to be 3°C below that of precise long duplexes, exactly as observed for both human and mandrill in Fig. 2*A*. The data are thus consistent with the presence in human and mandrill DNA of large numbers of *Alu* sequences identical to the precise consensus. There is no indication that the precisely hybridizing *Alu* sequences differ from the precise consensus in either species, but due to uncertainties they could differ 1% or so.

The interspecies divergence of the precise set of *Alu* sequences may be evaluated by comparing the human and mandrill melting curves of Fig. 2*A*. The fraction eluted at each temperature is almost exactly the same for the two species, showing that the precisely hybridizing *Alu* se-

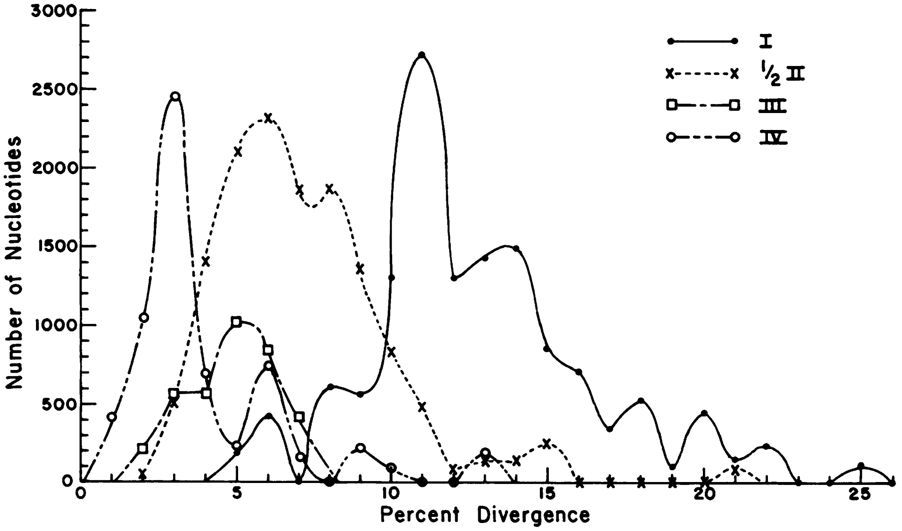

FIG. 1. Distribution of divergence of the *Alu* families from the precise consensus. The abscissa is divergence due to mismatches at non-CpG and nondiagnostic positions (5) as a percentage of such positions in an alignment with the precise consensus (5). All deletions and insertions are ignored. Plotted is the number of nucleotides in *Alu* sequences for intervals of 1% in divergence (thus allowing for the length differences of individual sequences). The classes (5, 8) are plotted separately. Class II is much more frequent than the others and the vertical scale was reduced by a factor of 2 for this class for easier comparison. About half of the *Alu* sequences in this set are >95% of full length, whereas 50 are less than half length. The 459 *Alu* sequences in GenBank were selected by J. Jurka on the basis of sequence similarity to the consensus. After removal of duplicates and artificially truncated sequences, 311 remained and 32 of these could not be classified, primarily due to large deletions, and are not included.

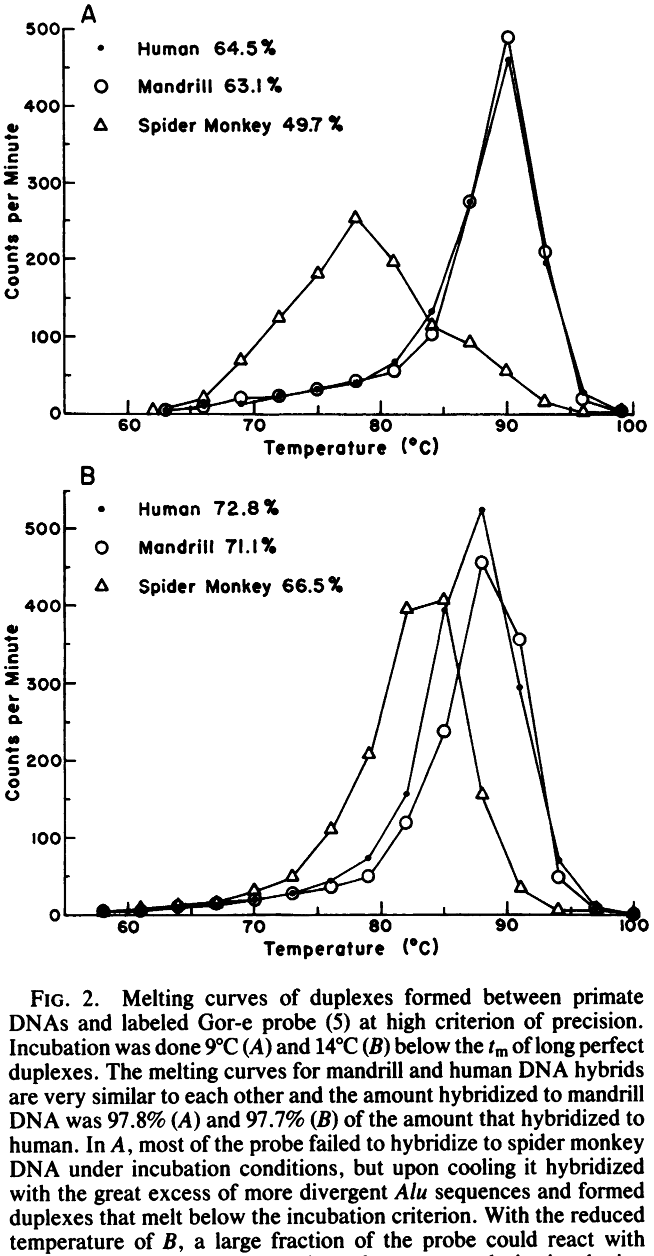

FIG. 2. Melting curves of duplexes formed between primate DNAs and labeled Gor-e probe (5) at high criterion of precision. Incubation was done 9°C (*A*) and 14°C (*B*) below the $t_m$ of long perfect duplexes. The melting curves for mandrill and human DNA hybrids are very similar to each other and the amount hybridized to mandrill DNA was 97.8% (*A*) and 97.7% (*B*) of the amount that hybridized to human. In *A*, most of the probe failed to hybridize to spider monkey DNA under incubation conditions, but upon cooling it hybridized with the great excess of more divergent *Alu* sequences and formed duplexes that melt below the incubation criterion. With the reduced temperature of *B*, a large fraction of the probe could react with somewhat divergent spider monkey *Alu* sequences during incubation and only a small fraction hybridized during cooling. Note that in the human and mandrill cases very little of the probe was left to hybridize with divergent *Alu* DNA during cooling. The native duplex of the probe (replicative form) elutes from hydroxyapatite under these conditions at 93°C (5).

quences of the two species melt over the same range of temperatures and hardly differ at all from each other. A difference of even 0.5% would have been evident in Fig. 2*A* as a one-half degree difference in $t_m$. As a result each of the temperature fractions in the peak would have differed by 20% or 30% between the two species, which would have been easily detectable. We conclude that the sequence difference between the two species for the precisely hybridizing *Alu* sequences (<1%) is much less than the average difference for single-copy DNA (about 6%, see below).

In a second measurement (Fig. 2*B*) where the incubation temperature was lower, as expected, the $t_m$s of both duplexes with the probe are lower since more divergent *Alu* sequences could hybridize with the probe. The data show that the $t_m$ of the duplexes with mandrill *Alu* sequences is slightly higher than with the human *Alu* sequences, suggesting that there are many precise *Alu* sequences in the mandrill genome. To assess the quantities of precise *Alu* sequences by reassociation kinetics, the same labeled probe was reassociated with sheared mandrill and human DNA under conditions permitting imprecise as well as precise duplexes (0.12 M PB, 70°C) for different times and assayed by binding to hydroxyapatite. The extents of hybridization and melting curves were identical for human and mandrill driver DNA at each point. For all points there were about two-thirds divergent duplexes and about one-third precise (melting at 88°C and above). The precise component showed a half reaction at about $C_0t$ = 0.01, implying that more than about 200,000 copies were present in each genome.

As a control, the single-copy DNA divergence was measured for the DNA preparations used in this work. We used standard hydroxyapatite thermal stability analysis (21) for hybrids of $^{32}$P-labeled single-copy human DNA with human, mandrill, and spider monkey (*Ateles geoffroyi*) sheared driver DNA. The results were a $t_m$ reduction of 5.6°C between mandrill and human and 8.2°C between spider monkey and human single-copy DNA, in agreement with the values in the literature (22).

## DISCUSSION

**Presence of Similar Source Genes in Mandrill and Man.** The evidence described above shows that there are large quantities of precise recently inserted (class IV) *Alu* sequences that are nearly identical in sequence in human and mandrill. However, interspecies comparisons of *Alu* sequences in homologous positions (including examples of class IV) show that *Alu* sequences once inserted evolve at about the same rate as single-copy DNA (10–12). The best explanation is as follows. At the time the primate lineages diverged leading to apes and Old World monkeys a source gene already existed with a sequence nearly identical to its current sequence; this source gene remained in the genome of both lineages from that period to the present; and it was responsible for the retroposition of many *Alu* sequences. The evidence shows that this precisely similar set is absent from spider monkey DNA, although many divergent *Alu* sequences are present in this species (Fig. 2 *A* and *B*). The absence from spider monkey could be due to loss of the current source gene in the New World lineage. The more likely alternative is that the current source gene had not yet evolved at the time of the branch to New World monkeys. Few of the precise class IV sequences are >5% divergent from the source, as shown in Fig. 1, suggesting that the current source was not active before the time of the branch to the New World monkey lineage.

**Conservation and Evolution of the Source Genes.** The long-term conservation of the source genes can be seen from their sequence relationship with the 7SL RNA of the signal recognition particle. The simplest model is that the *Alu* sequence arose from the 7SL in several stages, including deletion of a central region, duplication, and then additional deletions or insertions of short blocks of sequence. Thus the *Alu* sequence can be aligned with the terminal regions of the 7SL (duplicated). The result is that the earliest (class I) *Alu* sequence differs from the 7SL sequence in only 23 scattered positions and two gaps, with 214 positions matching out of 281. Presumably the differences are the changes that have occurred in the 7SL up to the present and in the *Alu* source sequences from sometime in the period of the mammalian radiation up to an early stage in primate evolution when class I was dominant. Since that time the *Alu* source sequence has evolved from class I to class IV and an additional 20 substitutions have occurred plus a few that later reverted to the 7SL nucleotide. Clearly the *Alu* source genes are conserved except during the periods of change from one class to another. We do not know whether each source gene that formed a class of *Alu* sequences existed as a single copy or cluster of similar genes.

Another aspect of the history of the *Alu* sequences and their source genes is worth comment. The estimate of the fraction of *Alu* sequences in the precise class from gene region sequence data can be made from Fig. 1 and agrees with the original estimate of the precise class as only a few percent of all of the *Alu* sequences (5, 6). However, the estimates in this paper and the previous work (5) both agree that the number of precise class *Alu* sequences is much larger, perhaps as much as 25% of all of the *Alu* sequences. This contrast suggests that at present there is selection against *Alu* sequence being inserted in gene regions. The total number of *Alu* sequences of all classes is about the same in gene regions as in the rest of the DNA (23). The implication is that *Alu* sequences may turn over in the nongene regions as expected since deletion due to unequal crossover between *Alu* sequences eliminates copies, but the rate cannot yet be estimated.

**What Are the Source Genes?** The *Alu* sequences can be thought of as a million pseudogenes that are by-products of source gene transcription. The conservation of the source gene implies that it is a functional gene that is under selectional constraint over its full length, owing to the role of its gene product. It is unknown what this role is or why so many pseudogenes should have been formed. We may guess that it is transcribed in germ-line cells since copies are so effectively incorporated in the genome and inherited. It is a good guess that it produces RNA for ribonucleoprotein particle since that is consistent with its small size, full-length conservation, and close relationship of its sequence to the sequence of the 7SL RNA (5, 24) of the signal recognition particle (25). There is no evidence for internal *pol* III promoter function *in vivo*, and we suppose that the source gene has a 5′ transcription control region, absent from inserted *Alu* sequences. The search for the source gene is difficult since there are probably >1000 perfect *Alu* copies inserted in the genome that have not even lost CpG dinucleotides. How many of these copies are transcribed is an open question but we assume that most have not been inserted into regions where transcription control sequences are appropriate. It might be worthwhile to search in the DNA for perfect copies of the precise consensus where the 5′-ward region has sequence similarity to the regulatory sequences associated with a 7SL gene. It is possible that an RNA transcript with the exact sequence of the modern consensus could be found if the appropriate tissue or cell types were examined.

**Interpretation of Conservation of Other Repeat Families.** In analogy with *Alu* source genes, the close interspecies similarity between other families of repeats could possibly be explained by the existence of an unknown gene that is conserved, shared between the species, and responsible for the production of more members of the family of repeats. Thus neither conservation of the sequences of all of the members of the families of repeats nor horizontal transfer is a necessary explanation. As a result, observations suggesting horizontal transfer of repeats or mobile elements are now more uncertain in their interpretation since conserved source genes will have to be considered as an alternative explanation. The retroposition mechanism has been identified as a source of small numbers of pseudogenes or low-frequency repeated sequences and now must be considered as a primary source of high-frequency repeated sequence families.

**Effect of *Alu* Sequences on the Primate Genome.** Deletion or duplication caused by the interspersed *Alu* repeats is a significant source of variation. Some major aspects of primate DNA sequence organization may be a result of balance between events of duplication and deletion leading to increased number of copies of some regions and loss of others. Thus many low-frequency duplicated and reduplicated regions probably exist in the majority of the DNA (which is not in gene regions) and many such copies date from the distant past and are quite divergent from each other. *Alu* sequences must have been inserted at least once per century in each lineage over the last 80 million years or so to account for nearly a million copies. Recently they have probably been inserted at a higher rate to account for the large number of precise (class IV) sequences, suggesting that some turnover may occur in nongenic regions. In comparison base substitutions are incorporated in the primate genomes at about 90 per generation or 450 per century. It is not known whether the number of *Alu* copies has increased steadily or in jumps or whether it has reached an approximate steady state with the insertion being compensated by deletions due to unequal crossover or other sorts of losses. There is evidence of interspecies differences in the number of *Alu* sequences among closely related primates (3).

**Effect of *Alu* Sequences on Primate Evolution.** *Alu* sequences affect the primate genome by their insertion and once inserted cause deletion or duplication of regions by unequal crossover. It appears that as much as 10% of cases of familial hypercholesterolemia are due to *Alu*-induced unequal crossover (26). Apparently more than half of Duchenne muscular dystrophy cases are due to deletions in this enormous gene (27). Deletions are a major source of most human genetic defects, but we do not know what fraction is due to unequal crossover at *Alu* repeats. The gross events of *Alu* sequence insertion or subsequent unequal crossover are clearly significant to human genetics but are not likely to cause the subtle changes in regulatory control systems that may be the prominent causes of evolutionary change. Nevertheless, *Alu* sequences may have caused duplications of genes, with potentially important effects.

The interspersed *Alu* sequences are apparently a by-product of the activity of the source genes, and the number of *Alu* sequences is the integrated effect of the source gene activity over many tens of millions of years in the past. The rate of deletion and duplication occurring in the primate genome in turn is partially or primarily due to the number of interspersed *Alu* sequences. Any change in the rate of introduction of *Alu* sequences would have a long delayed effect on the spacing of the *Alu* sequences and on the rate of deletion and duplication they cause. The resulting processes may have significant effects on variation and viability. The delay is so great that it is hard to believe that as a process of natural selection these potential defects or advantages could affect the source gene and in turn affect the number of interspersed *Alu* sequences. Thus the mechanisms are free of natural selection caused by the delayed effects of *Alu* interspersion. Further, since the *Alu* sequences can be produced by the source genes independent of the number of *Alu* sequences (though some copying may occur) the number of *Alu* sequences is not a primary cause of the production of more copies. It is hard to think of the *Alu* sequences as parasites or selfish genes if they are not responsible for most of their own production. The peculiarities of the family such as the interspersion of a million copies may be related to the lack of rapid "feedback" through natural selection. Such features of the genome that result from long period integration of a process could cause extinction or, if their contribution to variability, for example, was significant, they might be advantageous.

We thank Dr. G. Trabuchet for the gift of a clone of the *Alu* sequence recently inserted into gorilla DNA, Dr. J. Jurka for an alignment of GenBank *Alu* sequences with the precise consensus, Dr. O. Ryder for mandrill and spider monkey tissue, and Drs. E. Zuckerkandl, G. Dover, and E. Meyerowitz for criticism. This work was supported by National Institutes of Health Grant GM34031.

1. Waring, M. & Britten, R. J. (1966) *Science* **154**, 791–794.

2. Britten, R. J. & Kohne, D. E. (1968) *Science* **161**, 529–540.
3. Hwu, H. R., Roberts, J. W., Davidson, E. H. & Britten, R. J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3875–3879.
4. Weiner, A. M., Deininger, P. L. & Efstratiadis, A. (1986) *Annu. Rev. Biochem.* **55**, 631–661.
5. Britten, R. J., Baron, W. F., Stout, D. B. & Davidson, E. H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4770–4774.
6. Willard, C., Nguyen, H. T. & Schmid, C. W. (1987) *J. Mol. Evol.* **26**, 180–186.
7. Slagel, V., Flemington, E., Traina-Dorge, V., Bradshaw, H. & Deininger, P. (1987) *Mol. Biol. Evol.* **4**, 19–29.
8. Jurka, J. & Smith, T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4775–4778.
9. Quentin, Y. (1988) *J. Mol. Evol.* **27**, 194–202.
10. Sawada, I., Willard, C., Shen, C.-K. J., Chapman, B., Wilson, A. C. & Schmid, C. W. (1985) *J. Mol. Evol.* **22**, 316–322.
11. Maeda, N., Bliska, J. B. & Smithies, O. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5012–5016.
12. Miyamoto, M. M., Slightom, J. L. & Goodman, M. (1987) *Science* **238**, 369–373.
13. Britten, R. J. (1986) *Science* **231**, 1393–1398.
14. Coulondre, C., Miller, J. H., Farabaugh, P. J. & Gilbert, W. (1978) *Nature (London)* **274**, 775–780.
15. Trabuchet, G., Chebloune, Y., Savatier, P., Lachuer, J., Faure, C., Verdier, G. & Nigon, V. M. (1987) *J. Mol. Evol.* **25**, 288–291.
16. Economou-Pachnis, A. & Tsichlis, P. N. (1985) *Nucleic Acids Res.* **13**, 8379–8387.
17. Friezner Degen, S. J., Rajput, B. & Reich, E. (1986) *J. Biol. Chem.* **261**, 6972–6985.
18. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) *Cell* **15**, 1157–1174.
19. Deininger, P. L. & Slagel, V. K. (1988) *Mol. Cell. Biol.* **8**, 4566–4569.
20. Lin, C. S., Goldthwait, D. A. & Samols, D. (1988) *Cell* **54**, 153–159.
21. Hall, T. J., Grula, J. W., Davidson, E. H. & Britten, R. J. (1980) *J. Mol. Evol.* **16**, 95–110.
22. Benveniste, R. E. (1985) in *Molecular Evolutionary Genetics*, ed. MacIntyre, R. J. (Plenum, New York), pp. 359–417.
23. Moyzis, R. K., Torney, D. C., Meyne, J., Buckingham, J. M., Wu, J. R., Burks, C., Sirotkin, K. M. & Goad, W. B. (1989) *Genomics*, in press.
24. Ullu, E. & Tschudi, C. (1984) *Nature (London)* **312**, 171–172.
25. Blobel, W. (1982) *Nature (London)* **299**, 691–698.
26. Lehrman, M. A., Russell, D. W., Goldstein, J. L. & Brown, M. S. (1987) *J. Biol. Chem.* **262**, 3354–3361.
27. Chamberlain, J. S., Gibbs, R. A., Ranier, J. E., Nguyen, P. N. & Caskey, C. T. (1988) *Nucleic Acids Res.* **16**, 11141–11156.