MOLECULAR AND CELLULAR BIOLOGY, Oct. 1988, p. 4566–4569 Vol. 8, No. 10
0270-7306/88/104566-04$02.00/0

Recently Amplified *Alu* Family Members Share a Common Parental *Alu* Sequence

PRESCOTT L. DEININGER* AND VALERIE K. SLAGEL

Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, 1901 Perdido Street, New Orleans, Louisiana 70112

Received 18 April 1988/Accepted 28 June 1988

Three of the most recently inserted primate *Alu* family members are exceptionally closely related. Therefore, one, or a few, *Alu* family members are dominating the amplification process and the vast majority are not actively involved in retroposition. Although individual *Alu* family members are not under any apparent evolutionary constraint, the sequences of these active members are being moderately conserved.

The *Alu* family of repeated DNA sequences is one of the most successful of transposing elements, having duplicated itself approximately 500,000 times in the human genome (for reviews, see references 17 and 23 and P. L. Deininger, *In* M. Howe and D. Berg (ed.), *Mobile DNA*, in press). This duplication is thought to occur via an RNA intermediate in a process termed retroposition (15). Evolutionary studies have suggested that the *Alu* family began to amplify only about 60 to 70 million years ago (2, 4). However, recent studies of *Alu* family members in the primate globin genes suggest that relatively few of the *Alu* family members are the results of recent amplification events (9, 16). These data suggest that the *Alu* amplification rate may have been high at an early stage and have decreased recently.

There are three examples of *Alu* family amplification events that have occurred quite recently. One is an *Alu* family member adjacent to the β-globin gene region which is polymorphic in gorillas but not in the orthologous locations in chimpanzees or humans (19), suggesting an insertion less than 5 to 8 million years ago. Even more recent are two insertions of *Alu* family members which are present in some humans but not in others. Such insertions have been found in the *Mlvi-2* locus of 1 of 59 humans (6) and in one of two independent clones of the human tissue plasminogen activator gene (8). These last two insertions must have integrated less than about 200,000 to 1 million years ago (1).

All three of these recently inserted *Alu* family members fit within a subfamily (Fig. 1) that is thought to be younger than average *Alu* family members (18; C. Willard, H. T. Nguyen, and C. W. J. Schmid, J. Mol. Evol. in press). When all of these subfamily members were compared with one another in a pair-wise fashion (Table 1) and a tree of relationships was built (Fig. 2), using a least squares analysis (7), the close relationship of these recently inserted elements is striking. Thus, these three members are much more closely related to one another than they are to the other subfamily members, and they are even more distantly related to the average *Alu* family members.

The close relationship of the recently inserted *Alu* elements suggests they were either derived from a very small subset of closely related *Alu* family members or from a single *Alu* family member. In either case, the data suggest that the vast majority of *Alu* family members have little if any retroposition potential compared with that of this small subset. Individual *Alu* family members have been observed to evolve at about the rate of neutral evolution after their insertion (16). At about 0.5% per million years (10), this would suggest that the two human repeats diverged from a common ancestor about 4 million years ago and from the gorilla repeat about 8 million years ago. These numbers are slightly higher than those expected for the appearance of *Homo sapiens* (1) and its divergence from the gorilla (12). However, given the small sizes of the sequences for comparison, with only 5 to 12 differences between the sequence pairs, these data would still be consistent with the sequences having arisen from a single active progenitor.

This evolutionary analysis, showing that a very limited number of *Alu* family members may be capable of active retroposition, is consistent with a growing body of data indicating that expression of individual SINE members may be tightly restricted. It has now been demonstrated that the internal RNA polymerase III promoter may not always be sufficient for transcription in vivo. The human 7SL RNA gene, from which *Alu* is ancestrally derived (20), requires 37 bases of specific upstream sequence for transcription (21). Thus, since sequences upstream of the transcription unit cannot be retroposed, newly formed SINE members are likely to be inactive transcriptionally unless they integrate into an optimal chromosomal environment. This restriction of transcription of *Alu* is supported by the general lack of *Alu* transcription in HeLa cells (14) and by the predominant transcription of only one *Alu* family member in primate brains (11, 22). In a similar SINE family, the rat identifier family, it has been shown that a single gene codes for the major neuron-specific transcript, BC1 (3). These data demonstrate that the *Alu* family and SINE families are capable of very restricted, tissue-specific gene expression from individual members. Because there are additional steps after transcription in the retroposition process, it is possible that other factors (such as the variable A-rich 3′ end) could also contribute to the dominance of one, or a few, *Alu* family members in the amplification process.

There are two major evolutionary implications of these observations. The first is that, if a very small number of "active" retroposons can dominate the amplification process, mutations occurring in these members could easily have major affects on the process. Such mutations could result in the formation of newer "subfamilies" of altered sequences, as have been observed (18; Willard et al., in press), and also in changes in the retroposition rate of a family. Thus, it might not be unusual for variation to eventually "silence" a retroposon family. The second impli-

* Corresponding author.

```
              10        20        30        40        50        60        70        80        90       100
con   GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACGAGGTCAGGAGATCGAGACCATCCTGGCTAACACGG
tpa25 ................................................................................................C......A....
Mlvi  T...............................................................................................C......A....
gor   ...................A...............................................................................A
tpa1  ...T................A...A......................T.....A..T.............G..G.T.T.......G....A.C....T..
tpa14 A..G.....T.......................C..............A.....A.............................................
tkL   ....A.....A........T.............A..................T...T.A.........................................
tubE  ....A....T............T.....A...............................AG................A....T....C...........
tubF  ....A.....A..................x.........A.................A....TG......T............................A.
tubJ  A......T............A...............................T................................A..............
a2d1  ...........T..........x..............A....A....T.....T...A....GCx.......................TC.x....TG.A.
pjp   ...TA.......xx.T....................T..........T....A....T......T.................................T..
pro4  ...T......A...C.....C.........T.....G...................A.......A..........T.T......T..G......C....T..

              110       120       130       140       150       160       170       180       190       200
con   TGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCGTGGTGGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGCGT
tpa25 ......................C.........x...........A.....................TG.....T.........................A.
                               a
Mlvi  ........x...............................x...A.....x.....x.................T..........................
                            a
gor   ........................................................x...............G...........................
tpa1  ......................x..............T................A......A..T.........A.....................T..T.
                                    a
tpa14 ............................x.......................A.......................A..........A..T........A.
                                    a
tkL   ..................G.....................A.........T....A.........T............................G......
tubE  ...GG....................G...........T..AT.......T...T...........T...................................
tubF  ........T.........G.........................A...x....A...........................x.....C.xxxxxxx..CA..
tubJ  ......................T...................A.....x................T................x..................
                                                                                agcgactc
a2d1  ......T.T.......T.........A.G..........AC.AC.A....G.....x......x.......................G.............
                                            agcc
pjp   ..............................C..CC.....A.T.......T....A...................A.....................A.T.
pro4  ....................C.TA.A..............A..A.........ACA........A...T........A.........A.....G....C..T.

              210       220       230       240       250       260       270       280       290
con   GAACCCGGGAGGCGGAGCTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAA
tpa25 ....................................C.......................A.............................
Mlvi  ................xx..................C.....................................................
gor   ...........x..............................................................................
tpa1  .....Txxxx..T....G..................A.A...........T.................A......T...............
tpa14 ...................x.T..............A.A.......................................A...T....G....
tkL   ......A.......................T.....A.....................................................
tubE  ......A.....T......................................................T......TT...............
tubF  xxxxxxxxxxxxxxxxxxxxxxxxx....................T..........x.................Cxxx.AG..................
tubJ  ......A.........AGC...............C..T.T..........G..........A..........x.............T..T.
a2d1  .....T.C....T......................TxT...................TC.............................
                                      c
pjp   .....T.x....T.................A..............TCA.x..T.A....T.......A......................
                                                       caa
pro4  ......A..........G............T.....A.G...................A.GG...T.........T...T.G...G...
```

FIG. 1. Alignment of *Alu* subfamily members. The alignment of the sequences *tkL*, *tubE*, *tubF*, *tubJ*, *a2dl*, *pjp*, and *pro4* has been presented previously (18). The sequences *tpa25* (8), *Mlvi* (6), and *gor* (19) represent recently amplified *Alu* family members. The sequences *tpal* and *tral4* (8) were also added to the subfamily compilation. The consensus sequence (con) is that derived for these subfamily members and differs significantly from that for the overall *Alu* family (18). The dots indicate positions at which individual sequences agree with the consensus; sequence variations from the consensus are marked with the appropriate base, with x for deletions, or by insertions marked above the line.

cation is that the vast majority of the SINEs may simply represent pseudogenes of these active members. This would be consistent with previous models of selfish DNA (5, 13).

Nonetheless, the question remains of whether the active members themselves have a specific function. Our best estimate of the sequence of the typical active family member at any point in evolution comes from the consensus sequence for these family members. We have previously derived a consensus for the older *Alu* family members and the subfamily (18), and we now can generate a consensus for

TABLE 1. Divergence of *Alu* subfamily members

Sequence	% Divergence from:										
	tpa25	*Mlvi*	*gor*	*tpal4*	*tpal*	*tkL*	*tubE*	*tubF*	*tubJ*	*a2dl*	*pjp*
Mlvi	1.8										
gor	4.3	3.2									
tpal4	8.2	6.8	6.8								
tpal	13.2	11.4	11.4	12.8							
tkL	8.5	7.5	6.4	10.3	13.2						
tubE	10.7	9.6	9.3	13.2	15.7	11.0					
tubF	10.0	8.9	7.1	11.4	15.7	10.0	12.1				
tubJ	8.2	8.2	7.5	11.4	14.9	10.3	13.5	15.3			
a2dl	12.1	11.7	11.7	16.7	18.5	16.0	15.7	13.5	15.3		
pjp	12.8	12.5	11.4	14.6	14.2	13.2	15.3	14.6	19.6	17.1	
pro4	15.7	16.7	15.7	17.1	16.4	17.1	20.6	18.0	15.3	22.4	20.6

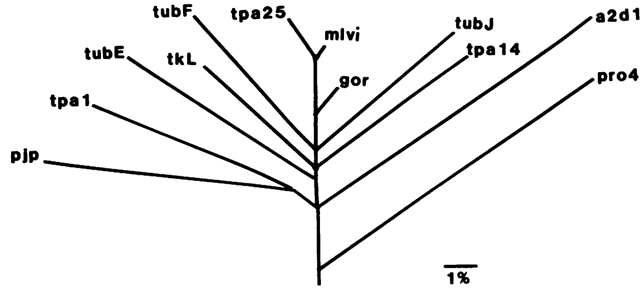

FIG. 2. Phylogenetic tree of *Alu* subfamily members. The pairwise comparisons of divergence, derived from point mutations, between all of the *Alu* family members shown in Fig. 1 is compiled in Table 1. These data were analyzed for relationships between the sequences by the program EVOLVE (7), which uses a least squares method of analysis. The best fit is presented and has a least squares fit of 7.6%. The lengths of the lines represent the relative divergences from a common ancestor, with the scale presented at the bottom.

these newly inserted *Alu* family members (Fig. 3). Analysis of these consensus sequences shows two interesting characteristics. First, the consensus for the new family diverges from the old consensus by only 5.3% (15 of 283 positions). This contrasts with the typical old *Alu* family member diverging from the same consensus by 14%. Since the typical *Alu* family member does not seem to be under significant selective pressures (9, 16), this suggests that active members are subject to a moderate degree of selection. Secondly, the *Alu* family consensus sequences are relatively rich in the dinucleotide CG. These CGs mutate rapidly in the individual *Alu* family members, and approximately two-thirds have been lost in the typical family member. Despite this normally high rate of mutagenesis for CG dinucleotides, there is very little alteration in CG nucleotides between the old consensus sequence and the sequences of the new *Alu* family members (Fig 3). Whether the selective pressure on active *Alu* sequences reflects an important function for *Alu* itself or simply the restraints placed on sequence by the retroposition process remains to be determined.

We thank S. Friezner Degen for a helpful discussion regarding the *tpa* gene sequence and C. Cohen, S. Squinto, M. Jazwinski, and J. Cook for critical readings of the manuscript.

This work was supported by Public Health Service grant GM29848 to P.L.D. from the National Institutes of Health; V.K.S. was supported by Public Health Service training grant CA09482 from the National Cancer Institute.

LITERATURE CITED

1. **Cann, R. L., M. Stoneking, and A. C. Wilson.** 1987. Mitochondrial DNA and human evolution. Nature (London) **325**:31–36.
2. **Daniels, G. R., M. Fox, D. Lowensteiner, C. Schmid, and P. L. Deininger.** 1983. Species-specific homogeneity of the primate Alu family of repeated DNA sequences. Nucleic Acids Res. **11**: 7579–7593.
3. **DeChiara, T. M., and J. Brosius.** 1987. Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. Proc. Natl. Acad. Sci. USA **84**:2624–2628.
4. **Deininger, P. L., and G. R. Daniels.** 1986. The recent evolution of mammalian repetitive DNA elements. Trends Genet. **2**:76–80.
5. **Doolittle, W. F., and C. Sapienza.** 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature (London) **284**:601–603.
6. **Economou-Pachnis, A., and P. N. Tsichlis.** 1985. Insertion of an Alu SINE in the human homologue of the Mlvi-2 locus. Nucleic Acids Res. **13**:8379–8387.

```
Old  GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACCTGAGGTCAGGA
                                                                     **
Sub  GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCAC--GAGGTCAGGA

New  GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCAC--GAGGTCAGGA

Old  GTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAA-TTAGCcGGGCGTGGTGGC
      *         *      *    *                                 *
Sub  GATCGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCGTGGTGGC
                   *      *                                                *
New  GATCGAGACCATCCCGGCTAAAACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCGTAGTGGC

Old  GCgcGCCTGTAgTCCCAGCTACTCGGGAGGCTGAGGCgGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTG
      *                                   *       *  *                  *
Sub  GGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCGGAGCTTGCAGTG
                            *
New  GGGCGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCGGAGCTTGCAGTG

Old  AGCCGAGATcGCGCCACTGCACTCCAGCCTGGGcgACAGAGCGAGACTCCGTCTC

Sub  AGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTC
               *
New  AGCCGAGATCCCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTC
```

FIG. 3. Comparison of *Alu* family consensus sequences. Three consensus sequences are presented, that for the bulk of *Alu* family members (old), that previously presented for the *Alu* subfamily (18; sub), and that derived from the three recently inserted *Alu* family members (new). The "old" consensus sequence has been slightly modified from that presented previously (18) to compensate for positions at which mutations occurring at CG dinucleotides had made the previous consensus somewhat ambiguous. The asterisks mark the positions of divergence of each of the consensus sequences relative to the subfamily consensus. All CG dinucleotide positions are underlined. Dashes mark the insertion and deletion of sequences that clearly distinguish subfamily members from older *Alu* family members.

7. **Fitch, W. M., and E. Margoliash.** 1987. The construction of phylogenetic trees: a generally applicable method utilizing estimates of the mutation distance obtained from cytochrome c sequences. Science **155**:279–284.
8. **Friezner Degen, S. J., B. Rajput, and E. Reich.** 1986. The human tissue plasminogen activator gene. J. Biol. Chem. **261**:6972–6985.
9. **Koop, B. F., M. M. Miyamoto, J. E. Embury, M. Goodman, J. Czelusniak, and J. L. Slightom.** 1986. Nucleotide sequence and evolution of the orangutan epsilon globin gene region and surrounding Alu repeats. J. Mol. Evol. **24**:94–102.
10. **Li, W.-H., T. Gojobor, and M. Nei.** 1981. Pseudogenes as a paradigm of neutral evolution. Nature (London) **292**:237–239.
11. **McKinnon, R. D., P. Danielson, M. A. Brow, F. E. Bloom, and J. G. Sutcliffe.** 1987. Expression of small cytoplasmic transcripts of the rat identifier element *in vivo* and in cultured cells. Mol. Cell. Biol. **7**:2148–2154.
12. **Miyamoto, M. M., J. L. Slightom, and M. Goodman.** 1987. Phylogenetic relations of humans and african apes from DNA sequences in the pseudo-n-globin region. Science **238**:369–373.
13. **Orgel, L. E., and F. H. C. Crick.** 1980. Selfish DNA: the ultimate parasite. Nature (London) **284**:604–607.
14. **Paulson, K. E., and C. W. Schmid.** 1986. Transcriptional inactivity of Alu repeats in HeLa cells. Nucleic Acids Res. **14**:6145–6158.
15. **Rogers, J.** 1983. Retroposons defined. Nature (London) **301**:460.
16. **Sawada, I., C. Willard, C.-K. J. Shen, B. Chapman, A. C. Wilson, and C. W. Schmid.** 1985. Evolution of the Alu family repeats since the divergence of human and chimpanzee. J. Mol. Evol. **22**:316–322.
17. **Schmid, C. W., and C.-K. J. Shen.** 1986. The evolution of interspersed repetitive DNA sequences in mammals and other vertebrates, p. 323–358. *In* R. J. MacIntyre (ed.), *Molecular evolutionary genetics*. Plenum Publishing Corp., New York.
18. **Slagel, V., E. Flemington, V. Traina-Dorge, H. Bradshaw, Jr., and P. L. Deininger.** 1987. Clustering and sub-family relationships of the Alu family in the human genome. Mol. Biol. Evol. **4**:19–29.
19. **Trabuchet, G., Y. Chebloune, P. Savatier, J. Lachuer, C. Faure, G. Verdier, and V. M. Nigon.** 1987. Recent insertion of an Alu sequence in the beta-globin gene cluster of the gorilla. J. Mol. Evol. **25**:288–291.
20. **Ullu, E., S. Murphy, and M. Melli.** 1982. Human 7S RNA consists of a 140 nucleotide middle repetitive sequence inserted in an Alu sequence. Cell **29**:195–202.
21. **Ullu, E., and A. M. Weiner.** 1985. Upstream sequences modulate the internal promoter of the human 7SL RNA gene. Nature (London) **318**:371–374.
22. **Watson, J. B., and J. G. Sutcliffe.** 1987. Primate brain-specific cytoplasmic transcript of the Alu repeat family. Mol. Cell. Biol. **7**:3324–3327.
23. **Weiner, A. M., P. L. Deininger, and A. Efstradiatis.** 1986. The reverse flow of genetic information: pseudogenes and transposable elements derived from nonviral cellular RNA. Annu. Rev. Biochem. **55**:631–661.