A master sequence related to a free left Alu monomer (FLAM) at the origin of the B1 family in rodent genomes

Yves Quentin
CNRS-LCB, UPR 7221, 31 Chemin Joseph Aiguier, 13402 Marseille, Cedex 20, France

Received April 8, 1994; Accepted May 16, 1994

ABSTRACT

The question of the origin of the B1 family of rodents is addressed. The modern B1 elements are similar to the left Alu monomer, but with a 9 bp deletion and a 29 bp duplication. Search of databases for B1 elements that do not exhibit those modern features revealed sequence fragments that are very similar to the free left Alu monomers (FLAMs) described in the primate genomes. In addition, the analysis reveals elements that have 10 bp or 7 bp deletion in place of the 9 bp deletion but without the 29 bp tandem duplication. The elements described define families of proto B1 elements (referred as PB1, PB1D10 and PB1D7) that appeared before the first modern B1 element. A phylogenetic reconstruction suggest that the origin of Alu and B1 families took place before the divergence between the primate and the rodent lineages and that each family has followed different evolutionary routes since this radiation.

INTRODUCTION

Most SINE (short interspersed elements) families are derived from known RNA polymerase III transcripts such as tRNAs or 7SL RNA (1–5), and it appears likely that they have been increased in number by retroposition, a process that involves a RNA intermediate (6,7). In theory, since the promoter for RNA polymerase III resides downstream of the transcription initiation site, each new element might be transcribed and prone to retroposition. Recently, this model has been refuted by statistical analysis of the retroposon sequences and more direct approaches (for review see ref. 8 and 9). Families of retroposons are mostly composed of silent copies, with a very small number of elements, the master sequences, able to duplicate by retroposition. Two evolutionary periods can be distinguished in the Alu family. The more recent period involves only dimeric master sequences that differ by single base changes (substitutions and insertion or deletion of one or two bases), and it is well documented (10 and references there in). The more ancient period includes the very first steps of the emergence of the Alu family and is characterized by large remodeling of monomeric sequences (11, 12).

A modern Alu element is about 300 bp long and is composed of two related sequences, the left and right monomers, arranged in tandem. This dimeric organization is a common feature in primates. The rodent B1 element is a monomer of about 140 bp (13–15), and contains an internal tandem repeat of 29 bp and a 9 bp deletion compared to the left Alu monomer (16, Figure 1). It has been suggested that the Alu and B1 elements arose from a 7SL RNA molecule (5, 17, 18). Sequence analysis has revealed free left Alu monomers (FLAMs) and free right Alu monomers (FRAMs) in primate genomes, which are older than the oldest Alu dimeric subfamily and are assumed to predate the first dimeric element (11, 19, 20). The FLAMs are composed of at least two subfamilies A and C (11). Each family or subfamily is characterized by a set of point mutations at diagnostic positions. The most noticeable characteristic is found in the FLAM-C master sequence, between positions 35 and 39 (throughout this paper, the numbering refers to the human 7SL RNA sequence (1–5). It corresponds either to three consecutive substitutions (TAC → ACT) in positions 36 to 38, or to a T deletion in position 35 with a T insertion between positions 38 and 39 (20). Since the B1 sequences have the same T deletion in position 35, we favor the second interpretation. Compared to the 7SL RNA sequence, the FLAM sequences have a deletion between positions 83 and 267, and the FRAM sequences have a smaller deletion between positions 97 and 239 and an additional 11 bp deletion between positions 247 and 257. Database screening for free Alu monomers that have neither the 11 bp deletion characteristic of the right monomers, nor the diagnostic base changes between positions 35 to 39 of the FLAM-C described above has revealed new monomeric elements referred as fossil Alu monomers (FAMs, 12). FAM descended from a 7SL RNA sequence but predated the FLAM and FRAM families (12). These monomeric families fill the gap between the parental 7SL RNA and the modern Alu elements. Since the B1 elements are also assumed to derived from the 7SL RNA, either both families appeared as the result of independent events directly from 7SL RNA sequences, or a single event is responsible for the Alu family in the primate genomes and B1 family in the rodent genomes. In order to test this alternative, we searched for fossil B1 elements in databases.The results obtained suggest that the B1 and Alu families have a common origin.

MATERIALS AND METHODS

Most of the sequence analyses were performed at the biocomputing server of Villejuif, France (for information send an electronic mail to bioinfo@genome.inserm-vjf.fr). Proto B1 elements were identified with the FASTA program and the significance of pairwise similarity score were estimated with the RSS program (FASTA package, 21, 22). Sequences were extracted from GenBank (23, Release 81.0) with the retrieval system ACNUC (24). The phylogenetic tree was reconstructed using the maximum-likelihood method (DNAML program of the PHYLIP package, 25).

RESULTS AND DISCUSSION

Compared to the 7SL RNA, and in addition to the central deletion, the modern B1 elements have a 29 bp tandem duplication and a 9 bp deletion between positions 65 and 73 (Figure 1). There are 5 substitutions between both halves of the duplication. The ancestral sequence of this duplication has been reconstructed by comparing the imperfect repeats of the B1-F master sequence (26), the 7SL RNA sequence (27), and the FLAM master sequence (11). This ancestral sequence is very close to the 7SL RNA and FLAM sequences and suffered only the A to T substitution in position 278 before the duplication. In the modern B1 elements, two substitutions occurred in the first repeat and three in the second repeat (Figure 1).

Description of proto-B1 elements

We assumed that the proto-B1 elements were similar to the FLAM described in primates (11, 20). Thus, we searched the rodent section of the databases for sequence fragments similar to FLAM-A master sequence (11). We used the complete master sequence as query, but better discrimination was obtained, between B1 elements and proto-B1 elements, with a shorter query sequence including the 9 bp deletion (see Figure 1).

```
                    10        20        30        40        50        60
H7SL        GCCGGGCGCGGTGGCGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCTGGAGGATCGC
FLAM-A      ..................C.................................G..........
B1-F        ........T........AC....T..A.......-...T.......A.....A..C.....T.
PB1D9       ........T........AC....T..A.......-...T.......A.....A..C.....T.

                 70        80               270       280       290
            _#########__________          __________________
H7SL        TTGAGTCCAGGAGTTCTGGG [183 bp] CCAGCCTGGGCAACATAGCGAGACCCCGTCTCT
FLAM-A      .....C..........GA.A          .................................
B1-F rep 1  .---------......GA..          .........T.T......
B1-F rep 2  .---------......CA..          A.....A....T...C..A..A....T.....A
PB1D9       .---------......GA..          ...........T......A..A....T.....A
                            @             @     @  @     @
```

Figure 1. Structural relationships between B1 elements and the 7SL RNA and FLAM sequences. Alignment of the human 7SL RNA sequence (H7SL: Reddy, 1988), FLAM master sequence (Quentin, 1992a), and B1-F master sequence (Quentin, 1989). B1-F rep 1 and rep 2 correspond to the 29 bp imperfect direct repeats depicted as overlined bases on the H7SL sequence. The diagnostic positions of rep 1 and rep 2 are marked by @ and the 9 bp deletion of the B1 element is signalled by #. PB1D9 sequence is the hypothetical master sequence that preceded the first modern B1 element characterized by the 29 bp deletion. The numbering refers to the H7SL sequence, and only the nucleotides that differ from the H7SL are listed. Notation: dot for nucleotide identity, dash for nucleotide deletion, and space for nucleotide insertion.

Since the proto-B1 sequences are very old, and since parts of the query sequence are found in B1 sequences, the FASTA outputs were contaminated by true B1 elements. In a second step, the true B1 elements have been discarded by means of sequence alignments with the B1 consensus sequence (26). 13 proto-B1 sequences have been selected with RSS, a program that gives a statistical estimation of the similarities observed between two sequences (21). Results of RSS show that, despite the low similarities observed between the FLAM master sequence and proto-B1 elements, the alignment scores can hardly have occurred by chance, and therefore confirm that reported elements found in rodents are related to the FLAM of primates and are members of a new B1 family: the PB1 family (Table 1).

Sequence alignments of the PB1 elements and of the FLAM consensus sequence are presented in Figure 2. The GenBank sequences MUSFSTC6 and MUSCYP345 start and end in the PB1 elements. The PB1 elements of MUSPNMT, RATMAD, and MUSNFIB6 are deleted in 5′ or 3′ ends. The other eight elements are full length PB1 sequences. All but two elements have an A-rich tail, and four of them are flanked by short direct repeats. MMATPB2 and RATATPB2S, corresponding to the 5′ end of the Na/K-ATPase beta 2 subunit gene, are probably orthologous sequences in the mouse and the rat genomes. A master sequence for the PB1 family can be derived from the sequence alignment. This sequence corresponds to the consensus sequence excepted for sites that have a high proportion of CpA, TpG, and CpG in the sequences and that are assumed to be CpG in the master sequence (28, 29). In the 5′ end of the sequence, ambiguous positions are signalled by question marks. All are CpGs in the FLAM master sequence (Figure 2). The PB1 master sequence have none of the diagnostic positions of the 29 bp tandem duplication, nor the substitution in position 278 that precedes the duplication.

The similarity search, performed with the FLAM master sequence as query, revealed two other kinds of proto-B1 elements. Indeed, among the sequence fragments extracted, some of them have a 10 bp or a 7 bp deletion in place of the 9 bp deletion observed in the B1 elements, and do not exhibit the 29 bp duplication. The 10 bp deletion is the one found in the 4.5S

Table 1. Features of the PB1 sequences extracted from GenBank.

GenBank name	positions	(bp)	strand	%sim[a]	score[b]	s.d.a.m.[c]	mean[d]	m.s.[e]
MMATPB2	184	294	-	0.67	165	18.75	38.5	63
RATATPB2S	972	1090	-	0.69	133	15.59	37.0	57
RATCTRPB-1	1672	1780	-	0.67	168	19.48	38.3	63
RATCTRPB-2	4247	4349	+	0.72	150	19.66	38.1	64
RNMYOLC1	881	994	+	0.65	176	26.34	35.9	50
RNTM4	4559	4667	+	0.73	174	22.03	42.7	59
MMIL5G	1694	1788	-	0.68	153	19.91	38.5	55
MUSOUAFRAA	3427	3540	-	0.67	169	27.35	35.0	53
MUSCYP345	8776	8865	+	0.71	162	19.30	38.7	57
MUSPNMT	1737	1820	+	0.75	173	19.51	39.9	64
MUSFSTC6	1	99	+	0.65	145	15.38	38.7	56
RATMAD	2174	2259	+	0.76	159	17.04	39.2	60
MUSFNIB6	1296	1219	-	0.67	132	17.86	35.	61

[a]similarity with the query (FLAM master sequence), [b]FASTA score, [c]standard deviation above mean, [d]mean FASTA score and [e]max score for 100 random shuffles of the sequence (results obtained with the RSS program, 21).

RNA sequence. This RNA is 90 to 94 bp long. It association with cytoplasmic poly(A+) RNA and its sequence conservation suggest that it could have some cellular function (15, 30–33). A specific search for elements that did not exhibit the 29 bp duplication confirmed the presence of 10 bp and 7 bp deleted elements (Figure 3). They constitute two new families of proto-B1 elements: the PB1D10 and PB1D7 families. Four subfamilies are recognizable in the PB1D10 family (Figure 3). The elements of the first subfamily (PB1D10-A) agree to the general consensus sequence of the PB1D10 family. The PB1D10-B and PB1D10-C subfamilies shared two diagnostic substitutions (positions 46 and 63), but PB1D10-B differentiates by C to T substitution in position 267. The PB1D10-D subfamily is defined by one diagnostic substitution (position 24). Compared to the Alu subfamilies, the PB1D10 subfamilies are based on few diagnostic positions, as observed with the B1 subfamilies (26). However, the primate element is twice as long as than its rodent equivalent. A master sequence had been derived for each family and subfamily (Figure 4). The 10 bp deletion cause a CpG doublet that can be accountable for the sequence variability observed at the edges of the deletion. The 5′ flank of the 7 bp deletion is also unusually mutated, and can be interpreted as the presence of two CpG doublets in the PB1D7 master sequence. The A to T substitution (in position 278), diagnostic of the sequence that precede the 29 bp duplication, is found in most of the PB1D10 and PB1D7 elements. However, the 5 diagnostic positions of the imperfect repeats are generally absents (Figure 4). Therefore, the PB1D10 and PB1D7 elements are not the result of homologous recombination between the repeats of B1 elements, but are proto-B1 sequences that appeared after the PB1 elements but before the modern B1 elements.

Sequences with a 9 bp deletion and without the 29 bp duplication have been found, but in all cases several diagnostic substitutions of the duplicated sequences are found, and so they are likely recombined B1 elements. Nevertheless, this result does not prove that PB1D9 elements never existed.

Phylogenetic relationship between proto-B1 sequences

The Figure 5 presents the evolutionary relationships between the 7SL RNA sequences, the 4.5S RNA, and the inferred master sequences of the monomeric families of primates and of the proto-B1 families and subfamilies described in rodents. The tree is based on the sequence alignment of Figure 4, from which we excluded positions 84 to 266, which correspond to the large deletions. The topology obtained is in agreement with the evolutionary steps that can be deduced from the analysis of the structural remodeling. The FAM is the first master sequence to appear as a result of a 141 bp deletion. Next is the bifurcation between the free right Alu monomers (11 bp deletion) and the other elements (42 bp deletion from the FAM). The master sequences of PB1 and FLAM-A are found very close to the node of this split. The next step is the separation between primates and rodents Alu-like families (10 bp and 7 bp deletion). On the rodent side, the diagnostic positions cannot resolve the order of appearance of the master sequences of the PB1D10 and, PB1D7 families. Nevertheless, the first master sequence of the modern B1 family arose from a sequence close to the PB1D10, PB1D7 master sequences that suffered a tandem duplication of 29 bp. The PB1D7 master sequence arose either from the PB1 master sequence throughout a 7 bp deletion and three consecutive substitutions or from the PB1D10 master sequence throughout 3 bp insertion. For reasons of parsimony, we prefer the second

```
                                    10        20        30                   40        50        60        70        80
FLAM-A                     GCCGGGCGCGGTGGCGCGCCTGTAGTCCCAGCTAC           TCGGGAGGCTGAGGC GGGAGGATCGCTTGAGCCCAGGAGTTC GAGA
PB1                        ......????......??..................          ............... A.......................... ...G
MMATPB2                    T..A...AT.C....-.ATAT...G.C.--.......         CT....TC...G..T AAA.....T.A.....GT......... A..G
RATATPB2S                  A..A...AT....A.-.ATAT...G.T.--.......TTGGGAG.T.....C...G..T AAA.....T.AC....GT.........AA..T
RATCTRPB-1                 ..TC.AT-TA.CA..AAAT.....C....A...A.GT         ..T.A.......A.. A....A..G...................- ----
RATCTRPB-2        TTCATTAAA..TAA..AGA....--.A.A...-..A....CAG...         .T.......AA.A.. ..C.----------.......T...... C..G
RNMYOLC1      GAATCTTAATCTGA.TA..TAT..A..-.TAT...CA..A.TT..A....         .TTCA..C......A ..T...G..A..G...T..T....... A..G
RNTM4      TGTGTTCCCCAGAGTGA.T....A.--...GC.T..G..TGC...--...AC.         CA.....C....... CCC......C..-...T..........G ..T.
MMIL5G         ATCAAAAGCTCT..T...T.T......A.AG......-.........G.         .----CTATCA.... AA...A...A.....A.........C.T ....
MUSOUAFRAA                 ..T....ATA....TT.A.A...A..A..T.T.T...         -A.A........... A.....-..A......T......TC.. A...
MUSCYP345                  -----.T.T.....TA.A.A.....GAG...T.TC..         ..A....A..--... A.....G..A..................C. A.C.
MUSPNMT                    ------------..A...G..........TT...G..         .A....C.........TT......C.A......G..G...A..G ..TG
MUSFSTC6                   ---------------..AT....A.....T...TG..         .T.....A..A.A.. AA....G..A.....A.......T.CG ...G
RATMAD                     -------------------------....A.....GT         ..CA...C.C....G AC..A...TC.G-......-......T ...G
MUSNFIB6                   ------------------------------------.         ..AA..A.A...... A......A.A.....A.ATT.A.C..- ...G

                              270       280       290
FLAM-A                     CC AGCCTGGGCAACATAGCGAGACCCCGTCTCT
PB1                        .. ..............................
MMATPB2                    T. .TT.A..A..G....A.A......TAA...AGAAGAAGAAAGGAGGAAGAGGAGGAGGAGGAAGA(GGA)10GGCAGCAGAAAGAAGA
RATATPB2S                  .A ..T.A..A..G....A.T......TAA...AGGAAAAAAAAAAAAAAAAAAAGGAGGAGGAAAGAAGAAAAGGAAGAAGA
RATCTRPB-1                 G. ..GT..A.....TGG.TA....GGT..T..AGGATAAAAGAAGAAGAAGCCAGCAACAACAATAAAAGAAAAA
RATCTRPB-2                 .. ....A.A..T.....ATA.C..A......TAAATTTTTTTATTAA
RNMYOLC1                   .. ...A..A.A..TG..AG.......T..TG.AAAAAAAAAAAAGAGGAGAAGCAAGGTGTGGTAAGGAATATCTTTAATCTCA
RNTM4                      .. ..T....C...TGC.-....G...T..G..ACAGATCAATAATTTTAAAAATCCAATGCAAATGAGTATGAGTTCCCTTGAG
MMIL5G                     ..T..AGA..T.CC...----------------CCCAGACACACACAACACCAGAAAGCT
MUSOUAFRAA                 .. .TTT.AAAT.G...CA.A...T.TT.G.CTCAAACCATTTTTTTTTTAATTAAAAATTTTTTTAAA
MUSCYP345                  TA ...T.T..GT..------------------
MUSPNMT                    A. T.A..A.A..--------------------
MUSFSTC6                   G. G..TCACTTG...GG..-....T.TT.TG.AAAATAAATCAAA
RATMAD                     .. .A...A.........AT.G.....--....ATTATTATT
MUSNFIB6                   .. .CAG....TG......TA.....T.AG...AACAAACGAACAAA
```

Figure 2. Alignment of the PB1 elements identified in GenBank with the FLAM master sequence. PB1 master sequence has been inferred from the alignment, and ambiguous positions are signalled by question marks. When they are distinguishable, the direct repeats that flank the PB1 elements were reported and underlined. The PB1 sequences are referred to by their names in GenBank.

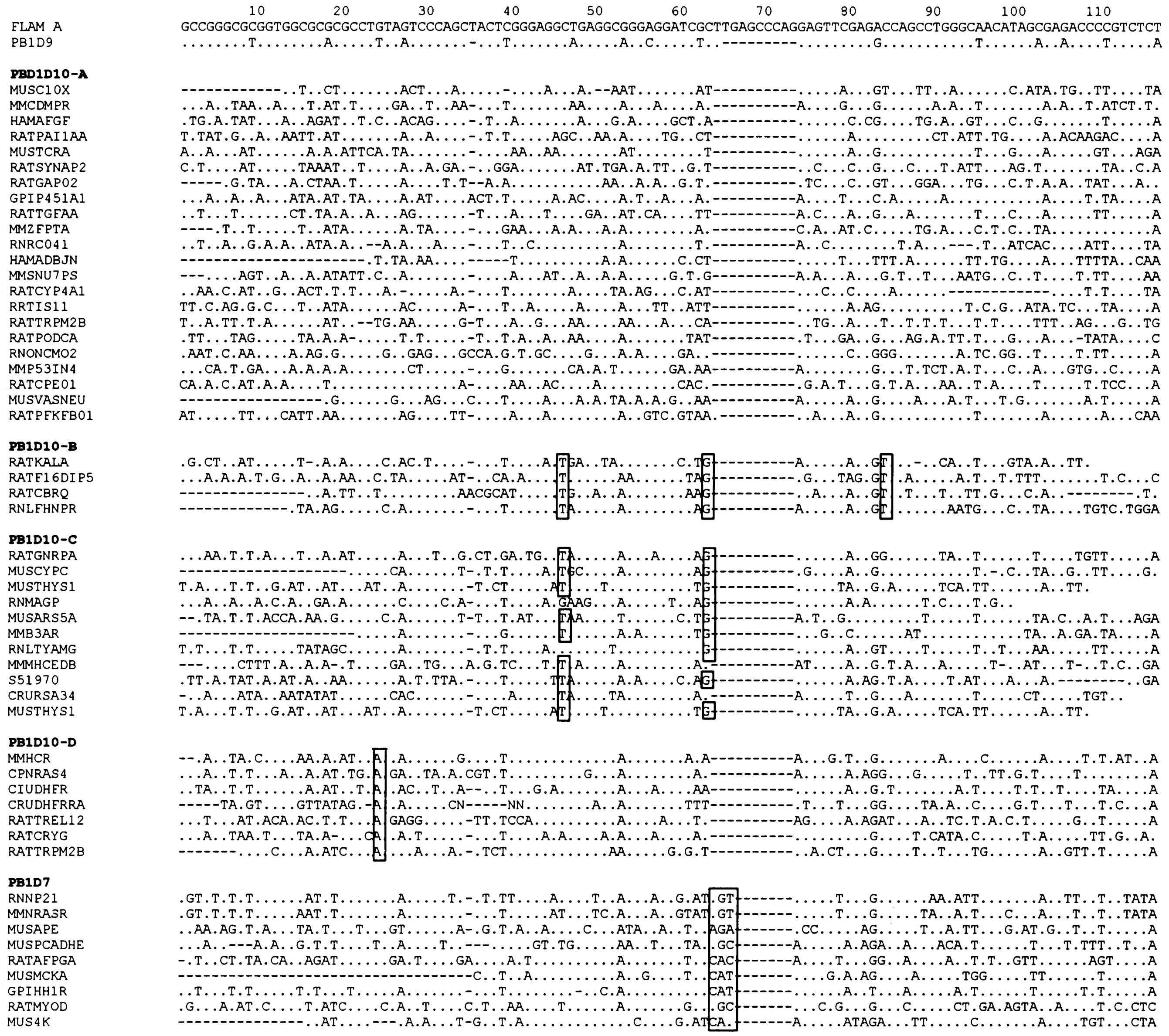

Figure 3. Sequence alignment of the PB1D10 and PB1D7 elements with the FLAM master sequence (same notation as Figure 1 and Figure 2). The diagnostic subtitutions of PB1D10 subfamilies are boxed.

hypothesis as it involves only one mutational event. The PB1D9 master sequence was either the product of a 9 bp deletion from the PB1 master sequence or the result of a T insertion in position 64. Since, the 4.5S RNA sequence and the PB1D9 hypothetical master sequence share the same T to A substitution in position 47, and since search for PB1D9 elements fail, it is tempting to speculate that the first modern B1 element arose from a PB1D10 master sequence throughout a single base insertion. Incidentally, a T insertion at that position increases the stability of the secondary structure of the PB1D9 RNA compared to the PB1D10 RNA, and then appears less fortuitous (9).

The average values of the pairwise similarity values computed between sequences of each family or subfamily are: 0.58 +/− 0.04 for the PB1 family, 0.56 +/− 0.05 for the PB1D10-A subfamily, 0.63 +/− 0.06 for the PB1D10-B subfamily, 0.64 +/− 0.05 for the PB1D10-C subfamily, 0.61 +/− 0.05 for the PB1D10-D subfamily and 0.61 +/− 0.06 for the PB1D7 family. All these values are lower than the one obtained with the older B1 subfamily (0.69 +/− 0.03, ref. 26), and are in agreement with the tree obtained, except for the PB1 family. Indeed, this later appears more conserved than expected. This deviation might be the result of the approach we used, since we were able to pick up only the more conserved members of the proto families. This is particularly true for the PB1 family, where the sequence region used as query in the FASTA searches is more conserved than its flanks (Figure 2).

The tree suggests that the FRAM master sequence differentiates before the divergence of primate and rodent lineages, but we did

```
                    10        20        30         40        50        60        70        80
H7SL        GCCGGGCGCGGTGGCGCGTGCCTGTAGTCCCAGCTACT-CGGGAGGCTGAGGCTGGAGGATCGCTTGAGTCCAGGAGTTCTGGGCTGTA
R7SL        .................AC...................-.............ACA..................................
FAM         ......????......??....................-..............?...............C..........GA.....C.
PB1         ......????......??C...................-..............A...............C..........GA..-----
R4.5S       .....TT.T........AC...G....[GATTTGCTGAA] ......A.....A........A.----------......GA..-----
PB1D10-A    ........T........AC.......A.......-...-..............A..........----------......GA..-----
PB1D10-B    ........T........AC.......A.......-...-.......T......A.........G----------......GA..-----
PB1D10-C    ........T........AC.......A.......-...-.......T......A.........G----------......GA..-----
PB1D10-D    ........T........AC....A..A.......-...-..............A..........----------......GA..-----
PB1D7       ........T........AC.......A.......-...-..............A..........CGC-------......GA..-----
PB1D9       ........T........AC....T..A.......-...-T.......A.....A..C.....T..---------......GA..-----
FLAM-A      ..................C...................-..............G...............C..........GA.A-----
FLAM-C     G...............T.AC.......A.......-...TT.......C.....G...............C..........GA.A-----
FRAM-A      ..................C...................G..............G...............C..........GA.....C.

           90                 240       250       260       270       280       290
H7SL       GTGCGCTAT  [141 bp] GATCGCGCCTGTGAATAGCCACTGCACTCCAGCCTGGGCAACATAGCGAGACCCCGTCTCT
R7SL       .........           .............................................................
FAM        ...??....           .............................................................
PB1        ---------           ----------------------------.................................
R4.5S      ---------           ----------------------------...........T...C.
PB1D10-A   ---------           ----------------------------...........T..............T......
PB1D10-B   ---------           ----------------------------T..........T..............T......
PB1D10-C   ---------           ----------------------------...........T..............T......
PB1D10-D   ---------           ----------------------------...........T..............T......
PB1D7      ---------           ----------------------------...........T..............T......
PB1D9      ---------           ----------------------------...........T......A..A....T.....A
FLAM-A     ---------           ----------------------------.................................
FLAM-C     ---------           ----------------------------.................................
FRAM-A     ...A.....           .........-----------...................G...G................A
```

Figure 4. Sequence alignment of the master sequences of the proto-Alu and proto-B1 families and subfamilies with the 7SL RNA and 4.5S RNA sequences. The GATTGCTGAA subsequences of the 4.5S RNA (Harada *et al.*, 1986) has no similarity with the other sequences.

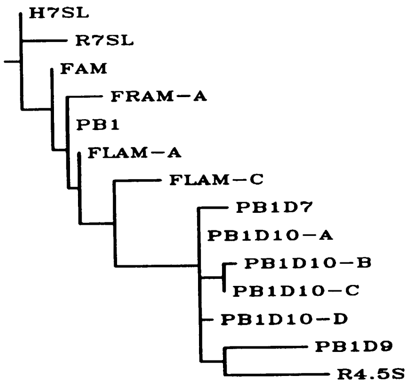

Figure 5. Evolutionary relationships between the proto-Alu and proto-B1 elements. The maximum-likelihood tree (Felsenstein, 1989) is based on the alignment of Figure 4, between positions 1 to 83 and 267 to 299. The lengths of the branches are proportional to the relative divergences from a common ancestor. All branches greater than 0 are in length significantly positive at the $p < 0.01$ level or $p < 0.05$ level for the branch that links PB1D9 and 4.5S RNA with the other sequences.

not find sequences related to the right Alu monomer in the rodent section of the databases. However, in the primate sequences, the FRAM sequences were also more difficult to identify than FLAM (11), because there were fewer copies. In addition, sequences evolved faster in rodents that in primates (34), so they are less similar to the query sequences when one search databases (the same applies to FAM elements). For example, the similarity values between the PB1 sequences (0.58) are far below the values obtained with the primate FAMs (0.68) Hence, one cannot conclude that sequences related to FRAM do not exist in rodent genomes; they may just have not yet been identified. Another explanation, suggested by the tree, is that the ancestor of the master sequences of the FRAM family was inactive for transposition before the divergence between primates and rodents and amplified only in the primate lineage.

CONCLUSION

We have reported the description of families of proto-B1 elements, referred as PB1, PB1D10, and PB1D7. The PB1 master sequence is collinear to the FLAM element of primates, and presents only two base substitutions compared to the FLAM-A master sequence. This observation suggest that FLAM-A and PB1 are lineage specific names for the same family of proto Alu/B1 elements that appeared before the mammalian radiation. Therefore, one can ask if members of this family are also present in other mammalian genomes. We did not find such evidence by database scanning, but a negative result cannot prove their absence. The analysis of orthologous sequences could be a more direct approach to answer such question.

Even if the modern Alu and B1 elements are now well differentiated, they have followed very similar evolutionary routes that can be separated into two periods. The first period initiated with the emergence of the first FAM master sequence in the ancestor of primates and rodents, and it is characterized by successive differentiations of master sequences throughout sequence remodeling of the internal part of the sequences (between position 64 and 266 in the 7SL RNA numbering). The separation between primates and rodents occurred after the amplification of the first PB1/FLAM-A elements. This period finished independently in each lineage by the appearance of the first master sequences of the modern Alu and B1 families. The modern Alu sequence was born from the fusion of a FLAM-C element with a FRAM element, and the modern B1 sequence results of a 29 bp tandem duplication in a PB1D10 or PB1D9 elements. The second period is characterized by successive waves of amplification, and a stabilization of the master sequences which evolved, between each wave, throughout substitutions and single base insertions/deletions.

ACKNOWLEDGEMENTS

We are grateful to G.Fichant and A.Cornish-Bowden for helpful comments and critical readings of the manuscript. This work is supported by the CNRS.

REFERENCES

1. Daniels, G. R. and Deininger, P. L. (1985) *Nature (London)*, **317**, 819–822.
2. Lawrence, C. B., McDonnel, D. P. and Ramsey, W. J. (1985) *Nucleic Acids Res.*, **13**, 4239–4252.
3. Rogers J. (1985) *Nature (London)*, **317**, 765.
4. Sakamoto, K. and Okada, N. (1985) *J. Mol. Evol.*, **22**, 134–140.
5. Ullu, E. and Tschudi, C. (1984). *Nature (London)*, **312**, 171–172.
6. Jagadeeswaran, P., Forget B. G. and Weissman, S. M. (1981) *Cell*, **26**, 141–142.
7. Van Arsdell, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. M., Manser T. and Gesteland, R. F. (1981) *Cell*, **26**, 11–17.
8. Schmid, C. W. and Maraia, R. J. (1992) *Curr. Opin. Genet. Dev.*, **12**, 874–882.
9. Quentin, Y. (1994) *Genetica*, in press.
10. Jurka, J. and Milosavljevic, A. (1991) *J. Mol. Evol.*, **32**, 105–121.
11. Quentin, Y. (1992) *Nucleic Acids Res.*, **20**, 487–493.
12. Quentin, Y. (1992) *Nucleic Acids Res.*, **20**, 3397–3401.
13. Krayev, A. S., Kramerov, D. A., Skryabin, K. G., Ryskov, A. P., Bayev A. A., and Georgiev, G. P. (1980) *Nucleic Acids Res.*, **8**, 1201–1215.
14. Krayev, A. S., Markusheva, T. V., Kramerov, D. A., Ryskov, A. P., Skryabin, K.G., Bayev A. A. and Georgiev, G. P. (1982) *Nucleic Acids Res.*, **10**, 7461–7475.
15. Haynes, S. R., Toomey, T. P., Leinwand, L. and Jelinek, W. R. (1981) *Mol. Cell. Biol.*, **1**, 573–583.
16. Rogers, J. (1985) *Int. Rev. Cytol.*, **93**, 187–279.
17. Ullu, E. and Weiner, A. M. (1984) *EMBO J.*, **3**, 3303–3310.
18. Ullu, E. and Weiner, A. M. (1985) *Nature (London)*, **318**, 371–374.
19. Quentin, Y. (1988) *J. Mol. Evol.*, **27**, 194–202.
20. Jurka, J. and Zuckerkandl, E. (1991) *J. Mol. Evol.*, **33**, 49–56.
21. Pearson, W. R. (1990) *Meth. Enzymol.*, **183**, 63–98.
22. Pearson, W. R. and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci.*, **85**, 2444–2448.
23. Burks, C., Cinkosky, M. J., Gilna, P., Hayden, J. E.-D., Abe, Y., Atencio, E. J., Barnhouse, S., Benton, D., Buenafe, C. A., Cumella, K. E., Davison, D. B., Emmert, D. B., Faulkner, M. J., Fickett, J. W., Fischer, W. M., Good, M. Horne, D. A., Houghton, F. K., Kelkar, P. M., Kelley, T. A., Kelly, M., King, M. A., Langan, B. J., Lauer, J. T., Lopez, N., Lynch, C., Lynch, J., Marchi, J. B., Marr, T. G., Martinez, F. A., McLeod, M. J., Medvick, P. A., Mishra, S. K., Moore, J., Munk, C. A., Mondragon, S. M., Nasseri, K. K., Nelson, D., Nelson, W., Nguyen, T., Reiss, G., Rice, J., Ryals, J., Salazar, M. D., Stelts, S. R., Trujillo, B. L., Tomlinson, L. J., Weiner, M. G., Welch, F.J ., Wiig, S. E., Yudin, K. and Zins, L. B. (1990) *Meth. Enzymol.*, **183**, 3–22.
24. Gouy, M., Gautier, C. and Milleret, F. (1985) *Biochimie*, **67**, 433–436.
25. Felsenstein, J. (1989) *Cladistics*, **5**, 164–166.
26. Quentin, Y. (1989) *J. Mol. Evol.*, **28**, 299–305.
27. Reddy, R. (1988) *Nucleic Acids Res.*, **16**, r71–r85.
28. Bird, A. P. (1980) *Nucleic Acids Res.*, **8**, 1499–1504.
29. Bird, A. P. (1986) *Nature (London)*, **321**: 209–213.
30. Harada, F., Kato, N. and Hoshino, H.-O. (1979) *Nucleic Acids Res.*, **7**, 909–917.
31. Harada, F. and Kato, N. (1980) *Nucleic Acids Res.*, **8**, 1273–1285.
32. Harada, F., Takeuchi, Y. and Kato, N. (1986) *Nucleic Acids Res.*, **14**, 1620–1642.
33. Jelinek, W. and Leinwand, L. (1978) *Cell*, **15**, 205–214.
34. Li W.-H. and Tanimura, M. (1987) *Nature (London)*, **326**: 93–96.