Volume 15 Number 10 1987 Nucleic Acids Research

# The L1 family (KpnI family) sequence near the 3' end of human β-globin gene may have been derived from an active L1 sequence

Atsushi Fujita, Masahira Hattori, Osamu Takenaka+ and Yoshiyuki Sakaki

Research Laboratory for Genetic Information, Kyushu University 18, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812 and +Primate Institute, Kyoto University, Inuyama, Aichi 484, Japan

Received February 27, 1987; Revised and Accepted April 24, 1987

## ABSTRACT

We previously reported that some L1 family (KpnI family) members are closely associated with the Alu family sequence. To understand the details of the L1-Alu association, the structure of a L1-Alu unit downstream from the β-globin gene was compared between human and primates. The results revealed that the L1-Alu-associated sequence was formed by the insertion of the L1 sequence, TβG41, into the 3' poly A tract of the preexisting Alu family sequence. It was estimated that the TβG41 sequence was inserted after the divergence of Old World monkeys and hominoids and before the divergence of orang-utan and common ancestor of other higher hominoids. From the calculation of the mutation rates of L1 sequences, it was suggested that the TβG41 was derived from an active L1 sequence which was able to encode reverse transcriptase-related protein.

## INTRODUCTION

L1 family, designated KpnI family previously, is a long interspersed repetitive DNA family which is present throughout marsupial and placental mammalian orders (1, 2). The L1 sequence is about 6 kb in full-length and repeated >10000 times per haploid genome (3). Several lines of evidences suggested that L1 family is a transposable element (4-7). Recent sequence studies showed that the primate and rodent L1 family sequences contain large open reading frames (ORFs) having significant sequence homology to some retroviral reverse transcriptases (8, 9). These findings together with the observation that the L1 family transcripts were detected in various kind of cells including pluriopotential embryonic cells (10-13) strongly supported the hypothesis that the L1 family was transposed by a mechanism including reverse transcription of the L1 transcripts. On the other hand, rearrangements including deletion, insertion

© IRL Press Limited, Oxford, England.

and inversion events have been often found in L1 family members (14, 15), and small circular DNAs containing L1 sequence have been identified (16, 17), which may be intermediates and/or by-product of such rearrangements. Careful analysis of these members which have undergone various genetic events may tell us some features of L1 sequences. We previously found curious association of L1 and Alu family sequences which were flanked by short direct repeats, and speculated that the L1-Alu associated unit might be a new mobile unit (18).

In this paper, we demonstrate that the L1-Alu association near the 3' end of the human β-globin gene may have been generated by the insertion of the L1 sequence, "TβG41", into the 3' poly A tract of the preexisting Alu sequence after the divergence of common ancestor of Old World monkeys and hominoids (about 26 million years ago) and before the divergence of orang-utan and common ancestor of the other hominoids except gibbons (about 13 million years ago). The results indicate that the L1-Alu-associated sequence is not a mobile unit but an association reflecting a tendency of retroposon to integrate adjacent to one another (4). Calculation based on the mutation rate of L1 sequences suggests that the TβG41 was derived from an active L1 sequence.

## MATERIALS AND METHODS

### Preparation of the genomic DNAs.

The species of primates examined include 10 species of Cercopithecoidea (Old World monkeys): *Macaca cyclopis* (Formosan monkey), *Macaca mulatta* (rhesus monkey), *Macaca assamensis* (Assamese monkey), *Macaca arctoides* (red-faced monkey), *Macaca nemestrina* (pig-tailed monkey), *Theropithecus gelada* (gelada), *Papio hamadryas* (Hamadryas baboon), *Erythrocebus patas* (patas monkey), *Cercopithecus aethiops sabaeus* (green monkey) and *Presbytis cristatus* (silvered lutong); and 5 species of Hominoidea (hominoids) : *Hylobates lar* (white-handed gibbon), *Pongo pygmaeus* (orang-utan), *Gorilla gorilla* (gorilla), *Pan troglodytes* (chimpanzee) and *Homo sapiens* (man).

Genomic DNAs except those of the Formosan monkey and the gorilla were prepared from lymphocytes of peripheral blood as described by Bell *et al*. (19). DNAs of the Formosan monkey and the gorilla were prepared from fetal thymus and lymph nodes, respectively.

Construction and screening of primate DNA libraries.

Partial *Eco*RI digests of primate DNAs were size-selected on sucrose density gradient and ligated into the arms of the phage vector Charon4A. The recombinant DNAs were packaged *in vitro* into phage particles and the resultant DNA library was screened by plaque hybridization method. The 750 bp *Eco*RI-*Pst*I fragment containing the 3' portion of human β-globin gene was used as a probe (20). DNAs from the positive clones were subcloned into pUC13 for further analysis.

Recombinant plasmids designated pRK12, pRK11, pRK19 and pRK20 were used for the analysis of human DNA (21).

DNA sequence analysis.

Cloned DNAs were digested by various restriction enzymes, and the resultant fragments were subcloned into pUC13 at appropriate sites. Some restriction fragments were digested with exonuclease BAL31 and were subcloned into *Sma*I site of pUC13. Nucleotide sequences of these cloned DNAs were determined by dideoxy chain-termination method as described by Hattori and Sakaki (22).

The nucleotide sequences were analyzed by GENAS system at Kyushu University Computer Center (23).

Southern hybridization analysis.

Five micrograms of DNAs were digested appropriate restriction endonucleases, electrophoresed in agarose gels, and transferred to nylon membrane filters. Probe DNAs were labelled with [$\beta$-$^{32}$P] dCTP by nick-translation. Hybridization was carried out in a solution containing 6XSSC (1XSSC:0.15M NaCl/0.015M sodium citrate), 10X Denhardt (1XDenhardt:0.02% bovine albumin/0.02% Ficoll-400/0.02% polyvinyl-pyrrolidone), 0.5% sodium dodecyl sulfate (SDS) and 0.5mg/ml salmon sperm DNA at 65°C. Filters were washed three times in a solution containing 2XSSC and 0.1% SDS at 65°C.

## RESULTS

### Absence of the TβG41-corresponding sequence in the Formosan monkey

An L1 family member, "TβG41", has been found downstream from human β-globin gene, which has a size of about 6 kb (20). In previous paper (18), we showed that the TβG41 is closely associated with *Alu* family sequence. In order to know the history of this L1-*Alu* association, we cloned and analyzed the corresponding chromosomal region of Formosan monkey. The 3' flanking region of the human β-globin gene except the TβG41 sequence was divided into 6 portions from A to F as shown in Fig.1, and DNA fragments from each portion were hybridized to the restricted fragments of the cloned Formosan monkey DNA. As schematically shown in Fig.1, the A portion was directly flanked to B in Formosan monkey. The results strongly suggested that a DNA sequence corresponding to the human TβG41 is missing in Formosan monkey.

To know the details of the structural difference demonstrated above, we determined the nucleotide sequence of the A-B junction region. Alignment of the nucleotide sequences of both species indicated that the *Alu* family associated with the TβG41 in the human genome was also present in Formosan monkey but the TβG41 sequence was missing (Fig. 2(a)). The results indicated that the TβG41 was inserted into the 3' adenine-rich region of the *Alu* sequence. The precise insertion site was, however, obscure because of the high adenine content of the region.

The 5' end of the L1 sequence was tentatively defined from the comparison of several "full-length" sequences (24), and the 5' end of the TβG41 was put at the position "1" shown in Fig.2(a) (25). However, our results suggested that the G-cluster flanking to the TβG41 was included in the transposed unit, although the size of the G-cluster was variable among hominoids (Fig.2(b)) for unknown reason. The genetic instability of the L1-*Alu* junction made it difficult to determine the precise 5' end of the transposed unit.

### The TβG41-corresponding sequences in various primate genomes.

Next, we asked when the TβG41 was generated during

Fig. 1. Restriction maps of human and Formosan monkey β-globin genes and their 3' flanking regions, and the summary of Southern hybridization analysis. The human TβG41 sequence is indicated by open bar. Following restriction sites are shown: EcoRI (E), BamHI (B), PstI (P), XbaI (X), BglII (Bg), EcoRV (EV), HindIII (H), HincII (Hc), SacI (S). DNA fragments from the portions, A-F, indicated at the top were used as probes for Southern hybridization analysis and portions of Formosan monkey which hybridized to each probe are indicated by open boxes (A-F) at the bottom. Heavy boxes, 1 and 2, are the probes used for Southern hybridization analysis of genomic DNAs shown in Table 1.

Fig. 2. Comparison of nucleotide sequences of the "TβG41" region of human and primates.
(a) The alignment of the human and Formosan monkey nucleotide sequences. Homologous sequences are shown by asterisks. Sequences in boxes are the direct repeats of the Alu family.
(b) The alignment of the 5' end of the "TβG41" sequence of man, gorilla and orang-utan. Homologous sequences among three species are shown by asterisks.

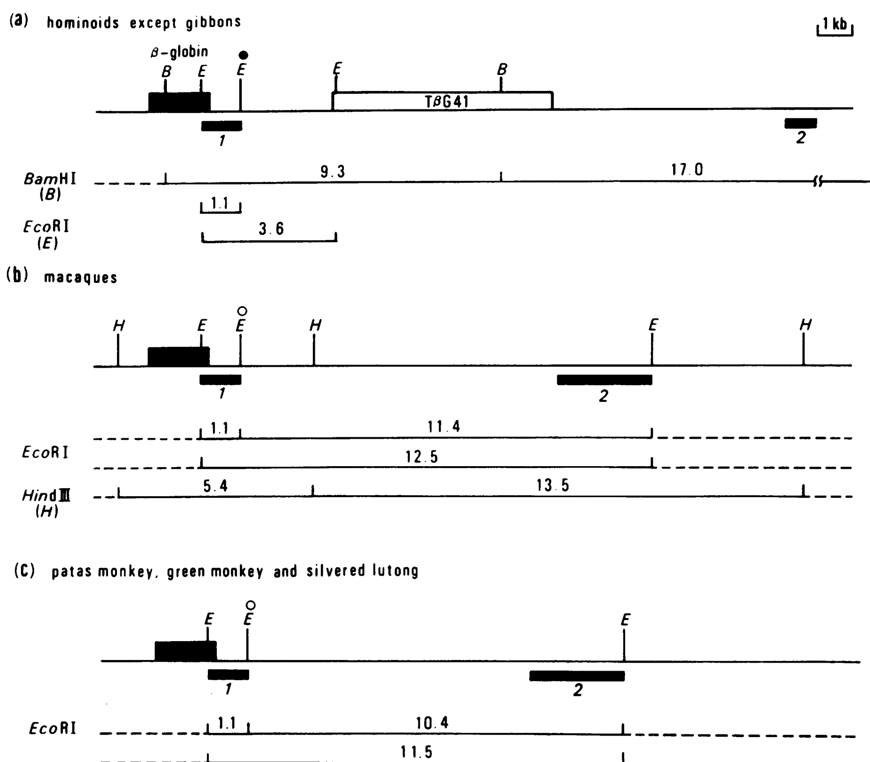


Fig. 3. Interpretation of the results shown in Table 1.
(a) The structure and restriction map of the 3' downstream of hominoid β-globin gene.
(b) The structure and restriction map of 3' downstream of macaque β-globin gene.
(c) The structure and restriction map of 3' downstream of patas monkey, green monkey and silvered lutong β-globin genes. The positions of the β-globin genes (heavy box) and the TβG41 (open box) are noted. And below the gene structure, the restriction maps of BamHI(B), HindIII (H) and EcoRI (E) are indicated. The EcoRI site marked (●) does not exist in human. The EcoRI sites that are very polymorphic are marked (○). The numbers above the lines are the lengths of fragments. The corresponding regions of probe 1 and 2 are also indicated below the lines (thick bars).

evolution. DNAs of various primates were digested with several restriction enzymes and subjected to Southern hybridization analysis. We employed two restriction fragments as probes, which do not contain repetitive sequences : probe 1, 1.1 kb EcoRI fragment which contains the 3' portion of the Formosan monkey β-globin gene; probe 2, 0.8 kb EcoRI-BglII fragment that

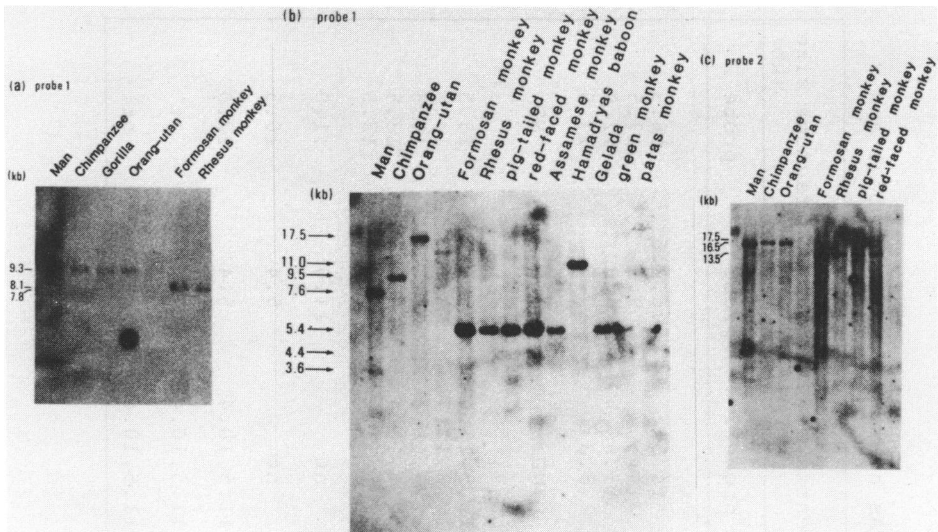


Fig. 4. Southern blotting profiles of various primate DNAs. Southern blotting experiments were carried out as described in MATERIALS AND METHODS. Species' names are indicated in the figure.
(a) DNAs digested by BamHI using probe 1.
(b) DNAs digested by HindIII using probe 1.
(c) DNAs digested by HindIII using probe 2.

is located about 7kb downstream from the human TβG41. Positions of these probe DNAs are shown in Fig.1. The results of the Southern blotting analysis are summarized in Table 1 and based on the restriction map of Fig.1, these results were interpreted as shown in Fig.3. The gorilla, the chimpanzee and the orang-utan showed hybridization patterns similar to those of human in BamHI digests (Fig.4(a)). These hominoids (and several Old World monkeys) showed 1.1kb hybridization bands in EcoRI digests which are absent in the human genome (Table 1). Barrie et al. (26) described the heterozygotes and homozygotes for this EcoRI fragment in gorilla. Thus, the 1.1 kb fragment may be due to restriction site polymorphism. Although HindIII digests gave different profiles from species to species for unknown reason (Fig.4(b), (c)), it is reasonable to consider that these hominoids have the "TβG41" sequences at the regions downstream from the β-globin genes. It is obscure whether the

Table 1 Sizes of the restricted fragments in primate DNAs hybridized with probe 1 and 2

| | *Bam*HI fragment size | | *Eco*RI fragment size | | *Hin*dIII fragment size (kb) | |
|---|---|---|---|---|---|---|
| species | probe 1 | probe 2 | probe 1 | probe 2 | probe 1 | probe 2 |
| hominoids | | | | | | |
| Man | 9.3 | 17.0 | 3.6 | 14.5 | 7.6,3.6 | 16.5 |
| chimpanzee | 9.3 | 17.0 | 3.6,1.1 | 14.5 | 9.5 | 17.5 |
| gorilla | 9.3 | 17.0 | 3.6,1.1 | 13.0 | 8.0 | 12.0 |
| orang-utan | 9.3 | 17.0 | 3.6,1.1 | 3.3 | 17.5 | 17.5 |
| white-handed gibbon | 14.0,4.3 | ND | 2.9 | ND | 13.5 | ND |
| Old World monkeys | | | | | | |
| Formosan monkey | 8.1 | 12.5 | 12.5,1.1 | 12.5,11.4 | 5.4 | 13.5 |
| rhesus monkey | 7.8 | 12.0 | 12.5,1.1 | 12.5,11.4 | 5.4 | 13.5 |
| Assamese monkey | 19.0 | 19.0 | 12.5,1.1 | 12.5,11.4 | 5.4 | 13.5 |
| red-faced monkey | 20.0,4.4 | 20.0,12.0 | 12.5,1.1 | 12.5,11.4 | 5.4 | 13.5 |
| pig-tailed monkey | 20.0 | 20.0 | 12.5,1.1 | 12.5,11.4 | 5.4 | 13.5 |
| Hamadryas baboon | 14.5,11.0 | 13.0,8.2 | 12.0,1.1 | nt | 11.0 | 11.0 |
| gelada | 14.5 | 13.0 | 12.0,1.1 | nt | 5.4 | 10.5 |
| green monkey | 15.0 | 15.0 | 11.5,1.1 | 11.5,10.4 | 5.4 | 22.0 |
| patas monkey | 17.0 | 17.0 | 11.5,1.1 | 11.5,10.4 | 5.4 | nt |
| silverd lutong | nt | nt | 11.5,1.1 | 11.5,10.4 | nt | nt |

ND : not detected , nt : not tested

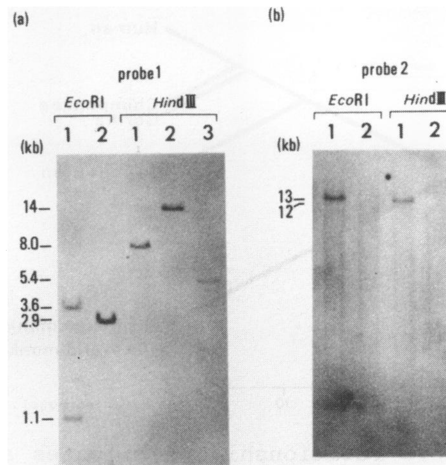


Fig. 5. Southern hybridization of gorilla, white-handed gibbon and Formosan monkey DNAs with probe 1 and 2. The numbers on the left side of each panel indicate the lengths (kb) of the detected fragments.
(a) Southern hybridization of gorilla (1), white-handed gibbon (2) and Formosan monkey (3) DNAs digested with EcoRI, and HindIII using probe 1.
(b) Southern hybridization of gorilla (1) and white-handed gibbon (2) DNAs digested with EcoRI and HindIII using probe 2.

white-handed gibbon has the "TβG41" or not, because probe 2 gave no hybridization band for the white-handed gibbon (Fig.5). The "probe 2" region may be missing in the white-handed gibbon.

The Formosan monkey and the other macaques showed the same blotting profiles each other in HindIII (Fig.4(b), (c)) and EcoRI digests, indicating that macaques have no "TβG41" sequence. The results of the other Old World monkeys were also interpreted as shown in Fig.3. It may be reasonable to consider that the green monkey, the patas monkey and the silvered lutong have no "TβG41" sequence as like macaques, because the distance between probe 1 and probe 2 regions in these monkeys were slightly shorter than that in Formosan monkey (Fig.3). Thus, we concluded that the "TβG41" sequence is absent in Old World monkeys, and that the "TβG41" sequence was inserted into the poly A tract at the 3' end of the Alu family sequence after the divergence of common ancestor of Old World monkeys and hominoids (about 26 million ago (27, 28)) and before the divergence of

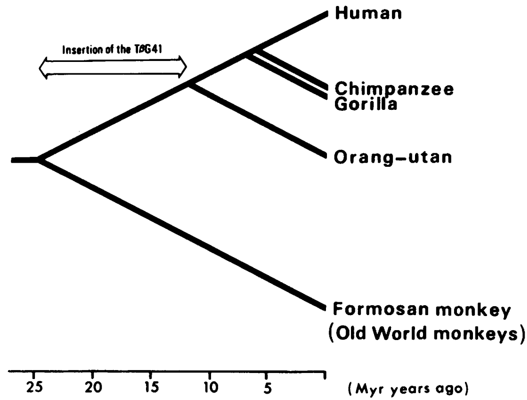


Fig. 6. Phylogenetic relationship of primates and the timing of the generation of the TβG41 sequence. We used the primate divergence times obtained by Hasegawa _et al_. (27, 28). The open arrow indicates the possible period for which the L1 family member (TβG41) may have been inserted into the _Alu_ sequence.

orang-utan and common ancestor of gorilla, chimpanzee and man (about 13 million ago (27, 28)) (Fig.6).

_The TβG41 may have been derived from an active L1 sequence._

Several evidences suggested that L1 sequence was derived from an active L1 sequence(s) which may be able to encode reverse transcriptase-like protein. Based on the mutation rates, we examined whether it is reasonable to think that the TβG41 was derived from such active L1 sequence.

Assuming that the L1 consensus sequence (9) represents an active L1 sequence at present, and that the TβG41 accumulated the mutations as a pseudogene, the timing when the TβG41 may have been derived from an active L1 sequence could be estimated as follows.

From the comparison of primate and prosimian consensus L1 sequences (9), we calculated the ratio of the evolutionary rates (k) from the evolutionary distances (K) of the "active L1 sequence" by the method of Kimura _et al_. (29, 30). We got the the evolutionary distances (K) for three codon positions ($K_1$=0.277, $K_2$=0.199, $K_3$=0.812) and synonymous component per third codon position (K's=0.637). Using the equation k=K/2T (T=time after divergency), we got the ratio of evolutionary rates (k); $k_1:k_2:k_3:k's$ = 0.44 (=$k''_1$):0.31 (=$k''_2$):1.28

($=k''_3$):1 ($=k''s$). In the consensus sequence of human L1 family, two ORFs were identified and the 2nd ORF (ORF2) encoded 1280 amino acid residues (9). From the comparison of the consensus ORF2 and the corresponding region of the human TβG41 (25), we found 128 point mutations in 113 codons, and 12 one base additions or deletions which cause frameshiftings. The percentage divergence ($D_{TL}$) between the consensus L1 and the TβG41 sequences is (128/3835) x 100 = 3.34 (%), which is the sum of the point mutations that occurred both in the TβG41 (dT) and the consensus L1 (dL) sequences. Thus we can divide the value ($D_{TL}$) into two components, dT and dL. Since the percentage divergence ($D_{TL}$) is relatively small in this case, we neglected the possibilities of double point mutations at one site. We also assumed that pseudogenes accumulate the mutations at the rate of 1.9 times higher than that of changes between synonymous codons (k's) (29, 31-33). Based on these data and assumptions, we got dT as follows.

$$dT:dL=1.9\times k''s:1/3\ (k''_1+k''_2+k''_3)$$

and

$$
\begin{aligned}
D_{TL} &= dT+dL \\
&= dT+1/3\ (k''_1+k''_2+k''_3)\times dT/(1.9\times k''s) \\
&= dT+1/3\ (0.44+0.31+1.28)\times dT/(1.9\times 1) \\
&= 1.356dT \\
&= 3.34\ (\%) \\
dT &= 2.46\ (\%)
\end{aligned}
$$

On the other hand, from the comparison of nucleotide sequences of the non-functional region near the 3' end of β-globin gene between human and Formosan monkey, we got the percentage divergence ($D_{HF}$=5.40%) (data not shown), which is the sum of mutations that occurred in the non-functional sequences of man and Formosan monkey. Assuming that Old World monkeys were diverged from man about 26 million years ago (27, 28), we estimated the timing (T) when the TβG41 sequence may have been derived from an active L1 sequence as follows.

$$dT:T=D_{HF}/2\ (\%):26\ (\text{million years})$$

$$
\begin{aligned}
T &= 2\times 26\times dT/DHF \\
&= 2\times 26\times 2.46/5.40 \\
&= 23.7\ (\text{million years})
\end{aligned}
$$

As described above, Southern hybridization analysis showed that the TβG41 was generated at 26~13 million years ago. Consequently it is likely that the TβG41 was derived from an active L1 sequence or gene.

DISCUSSION

We previously reported that the two L1 family members, TβG41 and HH10, are closely associated with Alu family sequence (11). Since the L1-Alu-associated sequences were flanked by short direct repeats, we speculated that they are new mobile elements. However, the present work showed that our previous speculation must be corrected. The previous results should be interpreted as follows: the direct repeats flanking to the L1-Alu-associated sequence were originally formed when the Alu family sequence was inserted into the chromosome and the L1 sequence was inserted at the 3' poly A tract of the Alu sequence afterward. Similar interpretation has been previously described by Rogers (4), who pointed out that retroposons have a tendency to integrate adjacent to one another. Our results provided an evidence for this hypothesis. It has been suggested that the integration site of L1 family appears to be AT rich (34), and it is the case in this L1- Alu association. AT rich sequence might be a hot spot of a certain recombinational event.

From the comparison of several L1 sequences, the 5' end of human L1 family sequences have been tentatively defined (24). The tentative 5' end of the TβG41 is indicated as position 1 in Fig.2. However as shown in Fig.2, it appears reasonable to consider that the G-cluster at the 5' end was included in the transposed unit. Thus, the previously-defind 5' end of the L1 sequence appears not to be definitive. Curiously the structure of the L1-Alu junction was variable among hominoids as shown in Fig.2(b). The molecular basis of this remarkable instability of this L1-Alu junction remains to be an open question.

An active L1 sequence(s) has been hypothesized, which possesses (or possessed) an ability to generate L1 family members. The analysis of the sequence divergency of the TβG41 suggested that it was derived from such active L1 sequence about

24 million years ago. The TβG41 may be new enough to tell us some characteristics of the active L1 sequence.

It was suggested that a rat L1 sequence has a potential role as a negative transcriptional regulatory element in rat insulin 1 gene regulation (35). However, the transcriptional levels of β-globin gene are indistinguishable between man and Old World monkey (36), so that the TβG41 sequence may not be involved in the transcriptional regulation of the β-globin gene.

ACKNOWLEDGEMENTS

We thank Dr. S. Kuhara for his advice on the computer analysis and Ms. H. Hamada and Y. Hayashi for assistance in preparation of the manuscript.

REFERENCES

1. Singer, M. F. and Skowronski, J. (1985) J. Trends Biochem. Sci. 10, 119-122.
2. Burton, F. H., Loeb, D. D., Voliva, C. F., Martin, S. L., Edgell, M. H. and Hutchison III, C. A. (1986) J. Mol. Biol. 187, 291-304.
3. Singer, M. F. (1982) Cell 28, 433-434.
4. Rogers, J. H. (1985) International Review of cytology 93, 87-279.
5. Voliva, C. F., Martin, S. L., Hutchison III, C. A. and Edgell, M. H. (1984) J. Mol. Biol. 178, 795-813.
6. Martin, S. L., Voliva, C. F., Burton, F. H. Edgell, M. H. and Hutchison III, C. A. (1984) Proc. Natl. Acad. Sci. USA 81, 2308-2312.
7. Katzir, N., Rechavi, G., Cohen, J. B., Unger, T., Simoni, F., Segal, S., Cohen, D. and Givol, D. (1985) Proc. Natl. Acad. Sci. USA 82, 1054-1058.
8. Loeb, D. D., Padgett, R. W., Hardies, S. C., Shehee W. R., Comer, M. B., Edgell, M. H. and Hutchison III, C. A. (1986) Mol. Cell. Biol. 6, 168-182.
9. Hattori, M., Kuhara, S., Takenaka, O. and Sakaki, Y. (1986) Nature 321, 625-628.
10. Skowronski, J. and Singer, M. F. (1985) Proc. Natl. Acad. Sci. USA 82, 6050-6054.
11. Lerman, M. I., Thayer, R. E. and Singer, M. F. (1983) Proc. Natl. Acad. Sci. USA 80, 3966-3970.
12. Giovanni, L., Haynes, S. R., Misra, R. and Jelinek, W. R. (1983) Proc. Natl. Acad. Sci. USA 80, 6533-6537.
13. Shafit-Zagardo, B., Brown, F. L., Zavodny, P. J. and Maio, J. J. (1983) Nature 304, 277-280.
14. Grimaldi, G. and Singer, M. F. (1983) Nucl. Acids Res. 11, 321-338.
15. Potter, S. S. (1984) Proc. Natl. Acad. Sci. USA 81, 1012-1016.

16. Jones, R. S. and Potter, S. S. (1985) Proc. Natl. Acad. Sci. USA 82, 1989-1993.
17. Schindler, C. W. and Rush, M. G. (1985) Nucl. Acids Res. 13, 8247-8258.
18. Miyake, T., Migita, K. and Sakaki, Y. (1983) Nucl. Acids Res. 11, 6837-6846.
19. Bell, G. I., Karam, J. H. and Rutter, W. J. (1981) Proc. Natl. Acad. Sci. USA 78, 5759-5763.
20. Adams, J. W., Kaufman, R. E., Kretschmer, P. J., Harrison, M. and Nienhuis, A. W. (1980) Nucl. Acids Res. 8, 6113-6128.
21. Shafit-Zagardo, B., Brown, F. L., Maio, J. J. and Adams, J. W. (1982) Gene 20, 397-407.
22. Hattori, M. and Sakaki, Y. (1986) Analytical Biochem. 152, 232-238.
23. Kuhara, S., Matsuo, F., Futamura, S., Fujita, A., Shinohara, T., Takagi, T. and Sakaki, Y. (1984) Nucl. Acids Res. 12, 89-99.
24. Grimaldi, G., Skowronski, J. and Singer, M. F. (1984) EMBO. J. 3, 1753-1759.
25. Hattori, M., Hidaka, S. and Sakaki, Y. (1985) Nucl. Acids Res. 13, 7813-7827.
26. Barrie, P. A. and Jeffreys, A. J. (1981) J. Mol. Biol. 149, 319-336.
27. Hasegawa, M., Kishino, H. and Yano, T. (1985) J. Mol. Evol. 22, 160-174.
28. Hasegawa, M., Kishino, H. and Yano, T. (1987) J. Mol. Evol. in press
29. Kimura, M. (1980) J. Mol. Evol. 16, 111-120.
30. Kimura, M. (1981) Proc. Natl. Acad. Sci. USA 78, 454-458.
31. Miyata, T. and Yasunaga, T. (1981) Proc. Natl. Acad. Sci. USA 78, 450-453.
32. Miyata, T. and Hayashida, H. (1981) Proc. Natl. Acad. Sci. USA 78, 5739-5741.
33. Li, W., Gojobori, T. and Nei, M. (1981) Nature 292, 237-239.
34. Furano, A. V., Somerville, C. C., Tsichlis, P. N. and D'Ambrosio, E. D. (1986) Nucl. Acids Res. 14, 3717-3727.
35. Laimins, L., Holmgren-Konig, M. and Khoury, G. (1986) Proc. Natl. Acad. Sci. USA 83, 3151-3155.
36. Martin, S. L., Vincent, K. A. and Wilson, A. C. (1983) J. Mol. Biol. 164, 513-528.