Selective cloning and sequence analysis of the human L1 (LINE-1) sequences which transposed in the relatively recent past

Hirohiko Hohjoh, Reiko Minakami and Yoshiyuki Sakaki
Research Laboratory for Genetic Information, Kyushu University 18, Fukuoka 812, Japan

Received May 8, 1990; Accepted June 13, 1990

EMBL accession nos X52230–X52235 (incl.)

ABSTRACT

L1 (LINE-1), a long interspersed repetitive DNA family of mammalian genomes, is thought to be a sequence family derived from a retrotransposon-like element(s), but its actively transposable unit(s) has not been identified yet. We developed a novel method for selective isolation of the human L1 sequences which transposed in a relatively recent past and may have still retained a feature of the 'active L1' unit. From the inspection of the nucleotide sequences, we conjectured that the 'active L1' or 'nearly acitve L1' units should have a high content of the CpG dinucleotide sequence, a mutation hot spot sequence, and contain several sites for rare cutters such as *BssH* II and *Nar* I at their 5′ terminal regions. Using these rare cutter sites as selection markers, the L1 sequences were isolated, which had the high content of CpG at the 5′ terminal regions and over 90% homology to L1 transcripts found in a human teratocarcinoma cell line. These L1s were shown to be 'relatively new L1' units which had integrated into chromosomes within these several million years during evolution. From the sequence data of these L1s and L1 cDNA, a consensus sequence of the 5′ terminal region of high CpG L1s were constructed. A region of the consensus sequence showed about 69% homology to the 5′ terminal region of *Drosophila jockey* element.

INTRODUCTION

L1 (LINE-1) family is a long interspersed repetitive sequence family which was initially found in human and primate genomes (1–6) and now throughout marsupial and mammalian orders (7–10). L1-like sequences have been also found in other organisms including insects (11–14), a protozoan (15) and a higher plant (16). L1 and L1-like sequences encode two large open reading frames (ORFs), one of which has a significant homology to reverse transcriptases of viral and retrotransposon origin (8, 11, 17, 18), strongly suggesting that L1 is a retrotransposon and has spread on the genome through RNA intermediates. The sequence characteristics of L1 such as the absence of long terminal repeat (LTR) and the presence of A-rich strtch at the 3′ end suggested that L1 belongs to a new class retrotransposon which may transpose by a mechanism different from those of typical retrotransposons. However, in mammalian including human, none of the actively transposable unit of L1 has been identified. In human (and primates), most of the L1 sequences so far analyzed were truncated at different positions of their 5′ ends and the remaining ones, although they were full-length units, were also disrupted by many mutations and rearrangements (see ref. 10, 19, 20 for review). It is conceivable that the recently transposed L1 sequences and their transcripts remain the features of transposable L1 units and give important imformation on L1 transposition. For example, L1 sequences (JH-27, JH-28) inserted to Factor VIII gene in a very recent past (21, 22), although truncated at their 5′ ends, showed that the actively transpoable L1 actually has a large ORF (previously proposed ORF 2) having a homology to reverse transcriptase. L1 transcripts of nearly full-length identified in a human teratocarcinoma cell line with embryonal carcinoma phenotype (NTera2D1) (23, 24) revealed that some L1s may be actively transcribed exactly from their 5′ ends under specific conditions. Thus, it is valuable to isolate the full L1 units that transposed in a recent past.

In this paper, we describe a strategy for selective cloning of the L1 genomic sequences that integrated into human genome in a relatively recent past and may have retained a feature of the 'active L1'. The L1 sequences isolated by this method show a high content of the CpG dinucleotide sequence at their 5′ terminal regions and over 90% homology to L1 transcripts found in NTera2D1 cells (24).

MATERIALS AND METHODS

Materials

Restriction endonucleases and other enzymes were obtained from Takara Shuzo (Kyoto, Japan) and Nippon Gene (Toyama, Japan). Radioactive compounds were purchased from Daiichi Pure Chemicals Co., Ltd. (Tokyo, Japan) and Amersham Corporation (England). Oligonucleotides were synthesized by ABI automated DNA synthesizer in our laboratory. Genomic DNAs of *Hylobates lar* (White-handed gibbon), *Pongo pygmaeus* (Orang-utan) and *Pan troglodytes* (Chimpanzee) were previously described by Fujita *et al.* (25). Human genomic DNAs were prepared from placenta as described previously (4).

Methods for recombinant DNA manipulation

Basic methods for manipulating recombinant DNAs such as Southern blotting, colony and plaque hybridizations, and transformation of *E.coli* were essentially the same as described by Maniatis *et al.* (26). Polymerase chain reaction (PCR) was carried out according to Saiki *et al.* (27) with some modifications. The nucleotide sequences were determined by dideoxy chain-termination method as described by Hattori and Sakaki (28) and analyzed by GENAS system at Kyushu University Computer Center (29).

The cloning of CpG rich L1s and their 5′ flanking sequences

Placenta DNA was digested with *Nar* I in complete and *Sau* 3A in partial. Alternatively, the DNA was digested with *Taq* I in complete. Digested DNAs were subjected to electrophoretical size fractionation on 1% agarose gels. The 0.5−3 kb fragment DNAs were electrophoretically eluted from the gels, ligated into the vector pUC 19, and transformed *E.coli* strain JM 83. Transformants were grown in L-broth and the plasmid DNAs were recovered as a whole. The plasmid DNAs were digested with *Bss*H II and subjected to the pulsed field polyacrylamide gel electrophoresis (PF-PAGE) to separate the digested DNAs from the undigested ones (30). The PF-PAGE was carried out on 4% polyacrylamide gels at 200V with 1 min of pulsed time for 16−20 hr at 10°C in a simplified hexagonal apparatus for CHEF (contour-clamped homogeneous electric field) gel electrophoresis as described by Ito and Sakaki (30). The linear DNAs were electrophoretically isolated from the gels, recircularized with T4 DNA ligase and transformed *E.coli* strain JM 83. The transformants were screened by colony hybridization using the *Bss*H II-*Nar* I fragment of L1 5′ terminal region (see RESULTS) as a probe.

RESULTS

A hypothetical 'active L1'

The 5′ terminal region of human L1 is G + C rich (10, 31, 32) but, as shown in Fig. 1, contains much less CpG dinucleotide sequences than the GpC dinucleotide sequences. We inferred that this feature reflected an evolutionary characteristics of L1 sequences. It has been known that the CpG dinucleotide sequence is a target site of DNA methylase on mammalian genomes and that the methylated cytosine residue is deaminated at a high frequency to give thymidine residue (33). Thus, the CpG dinucleotide sequence has a tendency to be lost at a high frequency on the genomes (33, 34). However, the regions surrounding the transcription start sites of many housekeeping genes and some tissue specific genes are known to be undermethylated and remain to be CpG rich (34, 35). These features of the CpG sequences led us to the idea that most of L1s have been inactivated by methylation and subsequent base substitution from CpG to TpG (or CpA), but the L1 units which retain their own transcriptional activities should have many CpG sites at their 5′ terminal regions. Thus, we considered that it may be possible to postulate 'active L1' sequence by converting all the TpG and CpA sequences in the human L1 consensus sequence (17) into CpG. The resultant hypothetical 'active L1' sequence was found to possess two rare cutter sites, *Bss*H II site at about 150 bp downstream from the 5′ end and *Nar* I site at further 300 bp downstream from the *Bss*H II site.

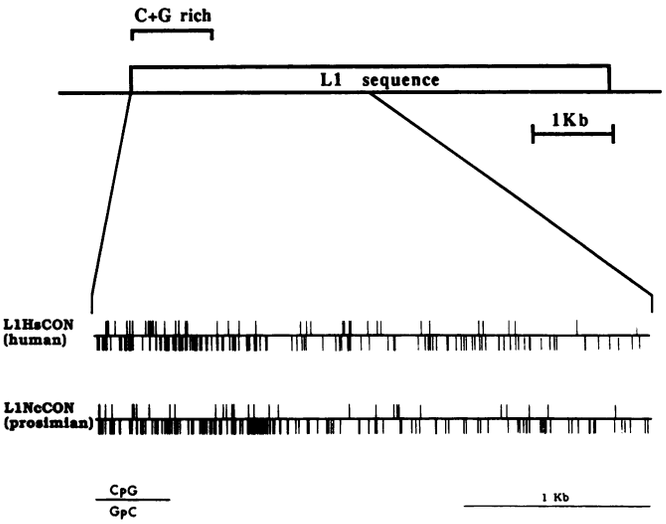

Fig. 1 The CpG and GpC maps in the 5′ regions of human and prosimian L1 consensus sequences. Based on the L1 consensus sequences by Hattori *et al.* (17), the CpG and GpC sites are marked with upbars and downbars, respectively.

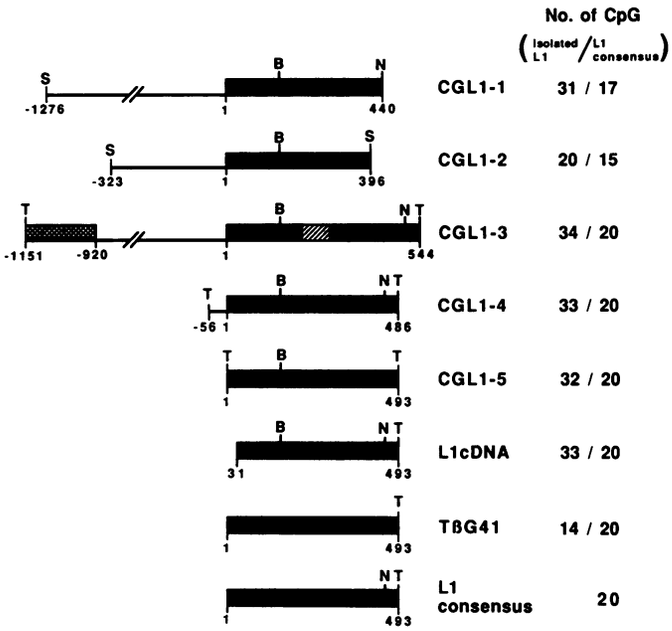

Fig.2 Schematic summary of the isolated CpG rich L1 sequences. The solid box indicates L1 sequence. The 5′ flanking region of L1 is indicated by horizontal thin line. The hatched box and the checkered box in CGL1-3 indicate the 39 nucleotide thymidine stretch and *Alu* sequence, respectively. L1 cDNA (24), TβG41 (31) and L1 consensus sequence (17) are also alined for comparison. The number of CpG sites in L1 sequences is shown in comparison with that in the corresponding region of consensus L1. B: *Bss*H II, N: *Nar* I, S: *Sau* 3A, T: *Taq* I

Selective cloning of 'relatively young L1s'

We asked whether the L1s containing both *Bss*H II and *Nar* I sites similar to the hypothetical 'active L1' were actually present on the human genome. Placenta DNA was digested by *Bss*H II and *Nar* I and subjected to Southern blot analysis. A very faint but detectable signal was observed at the position of approximate

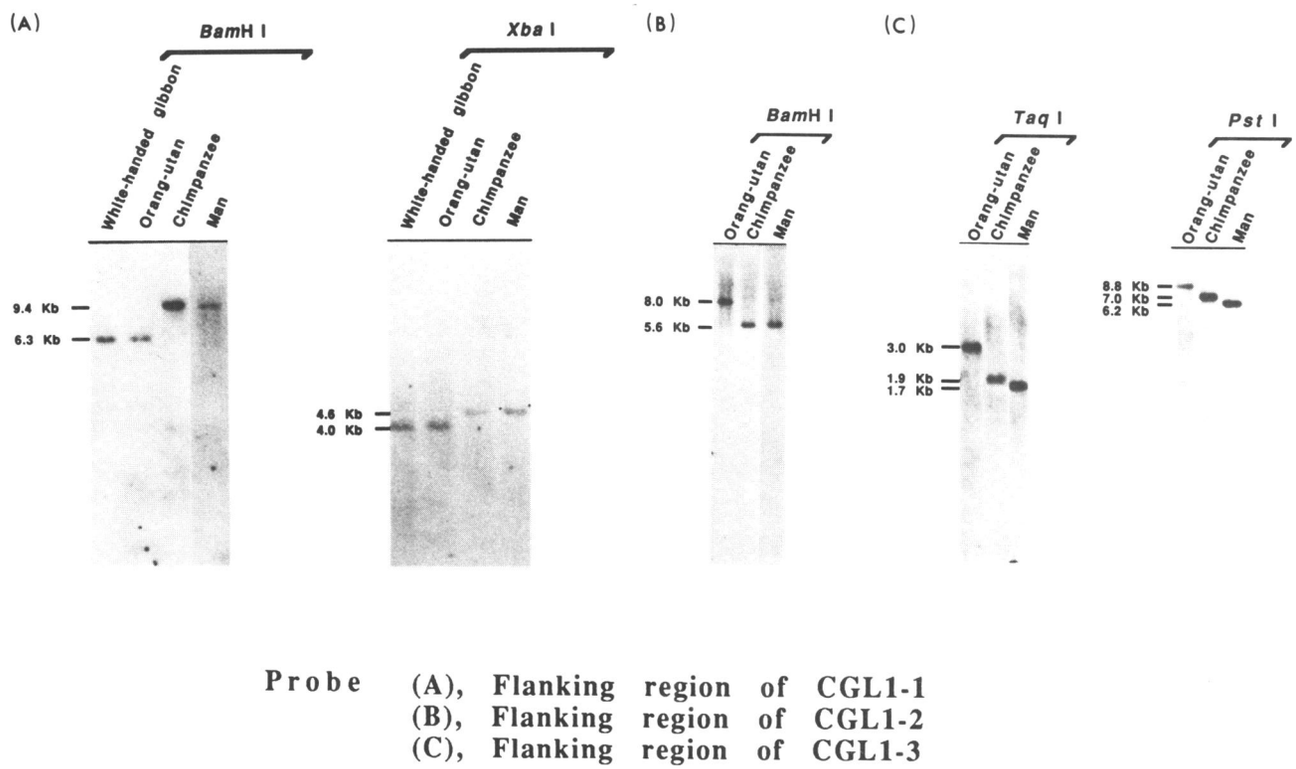

Fig. 3 Southern blotting profiles of various primate DNAs. Three micrograms genomic DNAs of the primates were digested with restriction endonucleases indicated, electrophoresed on 1% agarose gels, transfered to the filters, and hybridized with the 5′ flanking sequences of the isolated CpG rich L1s. The probe DNAs were prepared by polymerase chain reaction (PCR) (27) and labelled with [α-^{32}P] dCTP. For CGL1-3, the 5′ flanking region other than *Alu* sequence was used as a probe.

300 bp (data not shown). Cloning and sequencing study showed that the 300 bp *Bss*H II-*Nar* I fragments in the band were actually L1-derived, very CpG rich sequences (data not shown). These results indicated that the L1s having both *Bss*H II and *Nar* I sites were in fact present on the genome.

To isolate the CpG rich, *Bss*H II site-containing L1 units, a novel cloning strategy was developed. In principle, a human genomic DNA library was constructed using pUC plasmid vector and *Bss*H II site-containing L1 clones were selected by restriction enzyme digestion followed by colony hybridization. Details of the procedure is described in MATERIALS AND METHODS. A crucial step of the strategy was the separation of small number of plasmids linearized by *Bss*H II from the large mass of circular plasmids having no *Bss*H II site. The separation was successfully achieved by using pulsed field polyacrylamide gel electrophoresis (PF-PAGE) (30).

Using the procedure, the 5′ terminal and flanking regions of five *Bss*H II site-containing L1s were cloned. The structure of these L1 clones are shematically shown in Fig. 2. A notable feature of the present procedure is that it allows us to isolate the 5′ flanking regions of L1s. The flanking region may enable us to determine the 'age' (the time when the L1 was integrated into the chromosomes) and chromosomal location of the L1. At first, we estimated the age of these L1s by zoo blot analyses using their 5′ flanking sequences as probes (Fig. 3). In the case of CGL1-1 (Fig. 3-A), the white-handed gibbon and orang-utan showed 6.3 kb hybridizing bands in *Bam*H I digests, but the chimpanzee and the man showed 9.4 kb bands. In *Xba* I digests, 4.0 kb band was found in the white-handed gibbon and the orang-utan but 4.6 kb band in the chimpanzee and the man. The results suggested that CGL1-1 was generated after the divergence between orang-utan and the common ancestor of chimpanzee and man (about 13 million years ago) but before the divergence between chimpanzee and man (about 6 million years ago). Similarly, the integration of CGL1-2 was estimated to have occurred at a similar period (Fig. 3-B). On the other hand, CGL1-3 seemed to be integrated into the human genome after the divergence between chimpanzee and man (about 6 million years ago), because Southern blotting profiles of orang-utan, chimpanzee, and man were different one another (Fig. 3-C). The absence of CGL1-3 in chimpanzee genome was confirmed by cloning of the corresponding region of chimpanzee genomic DNA (data not shown). These selectively isolated L1s were thus relatively new in comparison with randamly isolated L1s such as TβG41 which was generated at 13−23 million years ago (25).

Characteristics of the 5′ terminal regions of the selectively isolated L1s

The 5′ terminal regions of these selectively isolated L1s were, as expected, shown to contain much more CpG sites than the human L1 consensus sequence (Fig. 2). The most CpG abundant L1 was CGL1-3 that contained 34 CpG sites in the region. The CGL1-3 included an insertion of a thymidine stretch flanked by 14 bp direct repeat sequences. The 5′ terminal regions of these CpG rich L1s are highly homologous one another and also have 90−95% homologies to L1 cDNA (cD11) isolated from a human teratocarcinoma cell line (NTera2D1) (24). We deduced the new L1 consensus sequence at the 5′ terminal region from the sequence data of these CpG rich L1s and L1 cDNA (Fig. 4). It is interesting to note that the 5′ flanking regions of these L1s

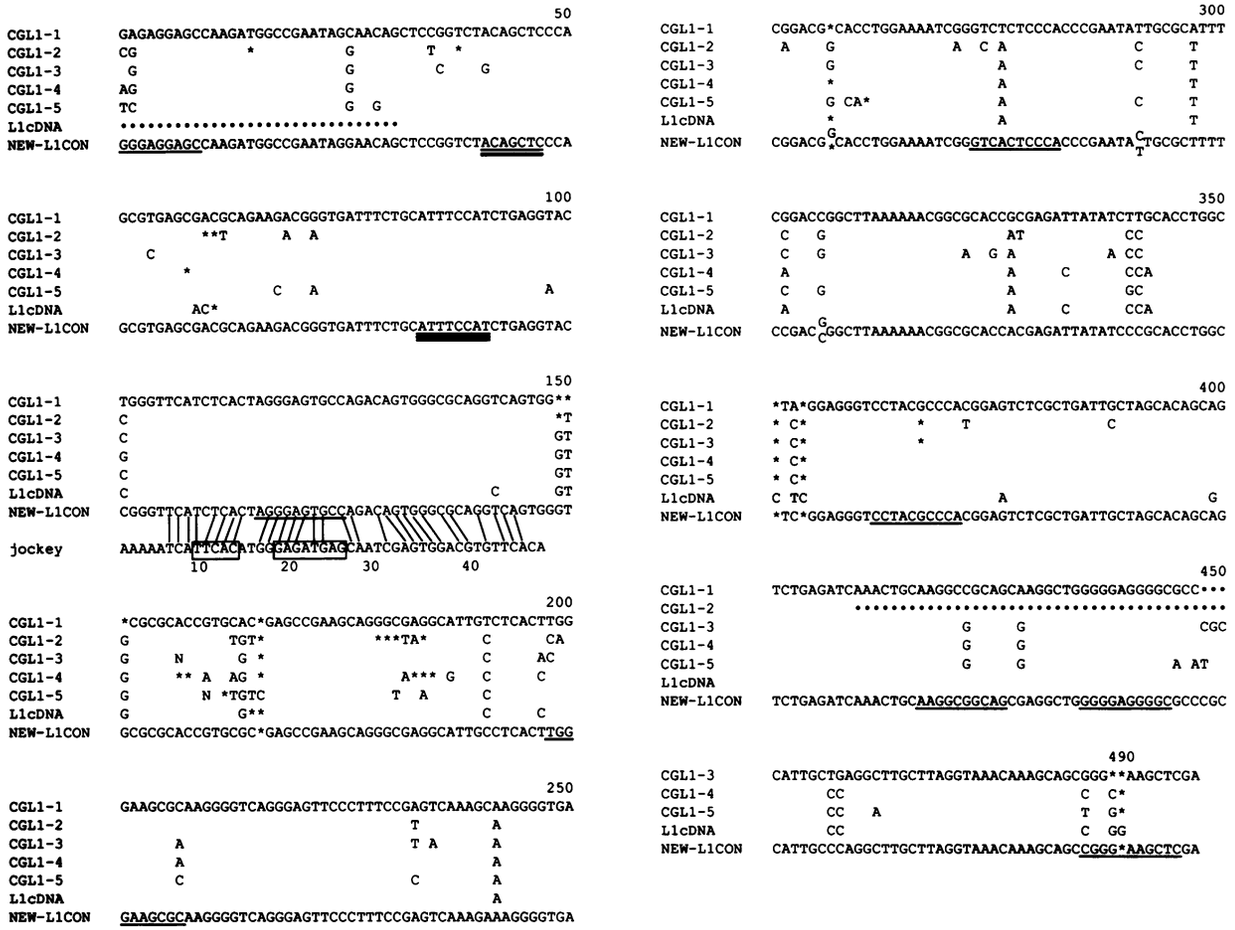

Fig. 4 Sequence characteristics of the selectively isolated L1s. Sequences of CGL1s and L1 cDNA are alined. Only nucleotides that are different from the sequences of CGL1-1 (position 1−447) and CGL1-3 (position 448−497) are indicated. Dotted lines indicate the regions lacking in the clones. Asterisk and N indicate one base deletion and undetermined base, respectively. The new L1 consensus sequence at the 5′ terminal region are constructed and shown at the bottom. Possible binding sites of Sp1 (38, 39) are underlined. Possible AP-4 binding site (40) and octamer motif (41) are indicated by double and triple underlines, respectively. The L1 consensus sequence was also compared with *jockey* element (13) and the homologous region is shown in the Figure. The core sequences identified by Mizrokhi *et al.* (37) are boxed. The accession numbers of these sequence data in the EMBL Data Library are as follows. CGL1-1 ; X52230 (1277, 1716), CGL1-2 ; X52231 (324, 719), CGL1-3 ; X52232 (1152, 1695), CGL1-4 ; X52233 (57, 542), CGL1-5 ; X52234 (1, 493).

were A + T rich (65−77% of A+T), supporting the idea that L1 has a tendency to be inserted at A + T rich regtion (36).

The initial event of L1 transposition may be its own transcription. Mizrokhi *et al.* (37) showed that *jockey, Drosophila* L1-like element, had the internal promoter for RNA polymerase II in the 5′ terminal region. Thus, the 5′ terminal region of L1 may also be of particular importance for the initiation of transposition. We searched for possible *cis*-acting regulatory sequences for transcription in the 5′ terminal region. Eight possible binding sites of a transcriptional factor SP-1 (38, 39) were found (Fig. 4). One of them at the nucleotide position of 5 was identical to the GC box of the P2 promoter of the rat insulin-like growth factor II gene (39). The sequences similar to AP-4 binding site (ACAGCTG) (40) and octamer motif (ATTTGCAT) (41) were also found at the positions of 44 and 87, respectively (Fig. 4). It should be noted that the region from the position 106 to 144 showed a significant sequence homology (69%) to the 5′ terminal region of *jockey* element which included the core sequence essential for the *jockey* internal promoter activity (37) (Fig. 4). The functional analysis of these elements is now in progress.

Structure of the full-length unit of CGL1-1

Only a part of L1 unit was isolated at the first step of the present procedure, but the 5′ flanking sequence may enable us to isolate the whole L1 unit. Using the 5′ flanking region of CGL1-1 as a probe, a genomic clone carrying a whole unit of CGL1-1 was isolated and shown to be a full-length L1 (6065 bp) having a nucleotide sequence homology of 97% to L1 cDNA (cD11) (24). From the predicted amino acid sequence of CGL1-1, two large ORFs of 338 a.a. (ORF 1) and 1274 a.a. (ORF 2), although interrupted by a few stop codons, were identified (Fig. 5-A). The two ORFs were in the same frame and separated by 39 bases including two conserved in-frame stop codons which were also present in the L1 cDNA (24). The ORF 2 of CGL1-1 had a homology of 94% at amino acid level to the ORF of JH-27 L1Hs which was recently inserted into the factor VIII gene of a hemophilia A patient (22). The 27 highly conserved amino acid residues in the reverse transcriptase region (42, 43) were identical between CGL1-1 and JH-27.

L1 sequences are classified into a few subsets by the characteristics of the 3′ trailer regions (44). The randomly selected genomic L1 units belong to the subset U. The full-length

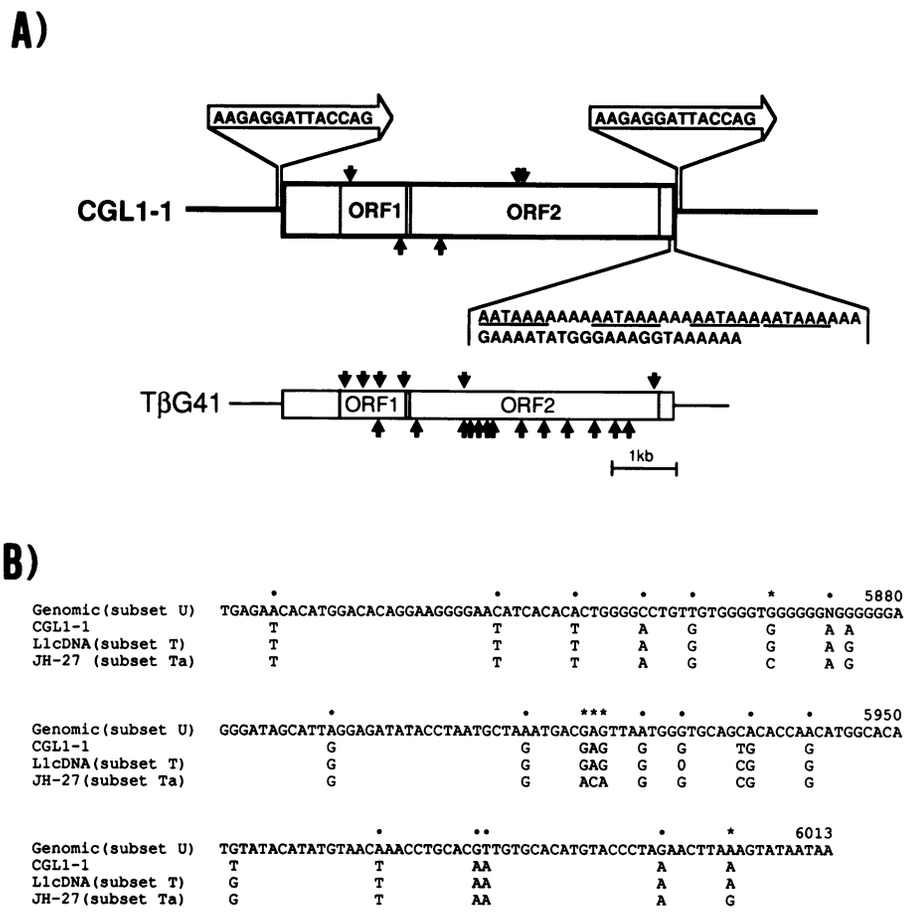

Fig. 5 Structure of CGL1-1. (A) Schematic drawing of the whole structure of CGL1-1. Downarrows and uparrows indicate stop codons and one-base deletions, respectively. The direct repeats at both ends of CGL1-1 are indicated by the open arrows in which the sequences are shown. The sequence at the 3′ end of CGL1-1 are shown and poly A signals are underlined. TβG41 (31) is shown as a comparison. (B) The sequences of the 3′ trailer regions. The 3′ trailer region sequences of genomic consensus L1 (subset U) (44), cDNA consensus (subset T) (44) and JH-27 (subset Ta) (21) are shown in comparison with CGL1-1. Dots show the positions at which nucleotides are different between subset U and T (Ta) (44), and asterisks ahow the positions at which subset Ta has its unique nucleotides (21, 22). The sequence data of the full-length CGL1-1 are registered with the accession number of X52235 in the EMBL Data Library.

L1 transcripts in a human teratocarcinoma cell line belong to the subset T. Some of the L1 transcripts and the L1 sequences transposed into the factor VIII gene, JH-27, JH-28 and JH-25, belong to a particular group of subset T, subset Ta (21, 22). Thus, subset T group, particularly subset Ta, seems to retain a feature of the 'active L1' unit. The CGL1-1 was shown to belong to the subset T in the 3′ trailer region (Fig. 5-B), suggesting that the CGL1-1 still retains the feature of 'active L1' unit to some extent.

DISCUSSION

From the findings that the 5′ terminal regions of L1s contain less CpG dinucleotide sequences than GpC and the fact that the CpG sequence is a mutation hot spot on mammalian genomes, we considered that the actively transcribed (and probably transposable) L1 units were CpG rich at their 5′ terminal regions. For selective isolation of such CpG rich L1s, a novel cloning procedure was developed by using a rare cutter site as a selective marker. Using the procedure, we have isolated the 5′ terminal regions of five L1 sequences. These L1s were, as expected, shown to be CpG rich at their 5′ terminal regions and to be transposed in a relatively recent past. They showed high sequence homologies to the L1 transcripts (L1 cDNA) in a human teratocarcinoma cell line with an embryonal phenotype (24) which may retain a feature of transcriptionally active L1s. On the basis of the sequence data of these selectively isolated L1s and L1 cDNA, we deduced a new L1 consensus sequence at the 5′ terminal region (Fig. 4). It should be noted that these 5′ terminal regions have no tandem repeat in contrast with the case of mouse and rat L1s (8, 45). The results suggest that the human L1 sequence, like *Drosophila* I, F, and *jockey* elements (11, 12, 37), may contain the signals necessary for transcription (and transposition) within the unit-length sequence. Interestingly, the consensus L1 contained the sequence resembling the core sites for the *jockey* transcription and also the possible binding sites for some transcriptional factors. It is interesting to see whether the 5′ terminal regions of L1s isolated in the present study have the promoter activity. Such study is now in progress.

The procedure developed in the present study enable us to isolate not only L1 5′ terminal sequences but also their 5′ flanking regions. As demonstrated above, the 5′ flanking regions can be used as probes to isolate the full-length L1 units as well as to estimate the age of the L1s. A full-length unit (CGL1-1) of high CpG L1 isolated in the present study was shown to have extremely high sequence homology to L1 cDNA and to be a

member of subset T to which L1 cDNAs and recently trasnsposed JH-27 belong. Thus, our method seems to be useful for selective isolation of the relatively 'young L1s' and these selectively isolated 'young L1s' may be valuable for understanding a feature of the 'active L1' units.

The insertion of L1 has irreversibly occurred at various locations on the chromosomes. Thus, L1s may be useful as a milestone of genome evolution, and selectively isolated 'young L1s' may provide us novel genetic markers for phylogenic studies and genome analyses of man.

ACKNOWLEDGMENTS

We would like to thank T. Ito (Nagasaki Univ.) for his advice on the PF-PAGE technique and for helpful discussion. We thank Dr. O. Takenaka (Kyoto Univ.) for supplying whole blood drawn from various kinds of primate and Miss H. Ohgusu for synthesis of oligonucleotides. We are grateful to Dr. H. Furuya and Miss K. Etoh for their advice on the computer search and Miss M. Mine for assistance in preparation of the manuscript.

REFERENCES

1. Adams, J. W., Kaufman, R. E., Kretschmer, P. J., Harrison, M., and Nienhuis, A. W. (1980) Nucl. Acids Res. *8*, 6113–6128.
2. Shafit-Zagardo, B., Maio, J. J., and Brown, F L. (1982) Nucl. Acids Res. *10*, 3175–3193.
3. Manuelidis, L., and Biro, P. A. (1982) Nucl. Acids Res. *10*, 3221–3239.
4. Sakaki, Y., Kurata, Y., Miyake, T., and Saigo, K. (1983) Gene *24*, 179–190.
5. Grimaldi, G., and Singer, M. F. (1983) Nucl. Acids Res. *11*, 321–338.
6. Lerman, M. I., Thayer, R. E., and Singer, M. F. (1983) Proc. Natl. Acad. Sci. USA *80*, 3966–3970.
7. 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.
8. Loeb, D. D., Padgett, R. W., Hardies, S. C., Shehee, W. R., Comer, M. B., Edgell, M. H., and Hutchinson III, C. A. (1986) Mol. Cell. Biol. *6*, 168–182.
9. D'Ambrosio, E., Waitzkin, S. D., Witney, F. R., Salemme, A., and Furano, A. V. (1986) Mol. Cell. Biol. *6*, 411–424.
10. Fanning, T. G., and Singer, M. F. (1987) Biochim. Biophys. Acta *910*, 203–212.
11. Fawcett, D. H., Lister, C. K., Kellett, E., and Finnegan, D. J. (1986) Cell *47*, 1007–1015.
12. Di Nocera, P. P., and Casari, G. (1987) Proc. Natl. Acad. Sci. USA *84*, 5843–5847.
13. Priimägi, A. F., Mizrokhi, L. J., and Ilyin, Y. V. (1988) Gene *70*, 253–262.
14. Xiong, Y., and Eickbush, T. H. (1988) Cell *55*, 235–246.
15. Kimmel, B. E., ole-Moiyoi, O. K., and Young, J. R. (1987) Mol. Cell. Biol. *7*, 1465–1475.
16. Schwarz-Sommer, Z., Leclercq, L., GÜbel, E., and Saedler, H. (1987) EMBO J. *6*, 3873–3880.
17. Hattori, H., Kuhara, S., Takenaka, O., and Sakaki, Y. (1986) Nature *321*, 625–628.
18. Doolittle, R. F., Feng, D. -F., Johnson, M. S., and McClure, M. A. (1989) Quart. Rev. Biol. *64*, 1–30.
19. Sakaki, Y. (1987) Mol. Biol. Med. *4*, 193–197.
20. Singer, M. F., and Skowronski, J. (1985) Trends Biochem. Sci. *10*, 119–122.
21. Kazazian Jr., H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., and Antonarakis, S. E. (1988) Nature *332*, 164–166.
22. Woods-Samuels, P., Wong, C., Mathias, S. L., Scott, A. F., Kazazian, Jr., H. H., and Antonarakis, S. E. (1989) Genomics *4*, 290–296.
23. Skowronski, J., and Singer, M. F. (1985) Proc. Natl. Acad. Sci. USA *82*, 6050–6054.
24. Skowronski, J., Fanning, T. G., and Singer, M. F. (1988) Mol. Cell. Biol. *8*, 1385–1397.
25. Fujita, A., Hattori, M., Takenaka, O., and Sakaki, Y. (1987) Nucl. Acids Res. *15*, 4007–4020.
26. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
27. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) Science *239*, 487–491.
28. Hattori, M., and Sakaki, Y. (1986) Analytical Biochem. *152*, 232–238.
29. Kuhara, S., Matsuo, F., Futamura, S., Fujita, A., Shinohara, T., Takagi, T., and Sakaki, Y. (1984) Nucl. Acids Res. *12*, 89–99.
30. Ito, T., and Sakaki, Y. (1988) Nucl. Acids Res. *16*, 9177–9184.
31. Hattori, M., Hidaka, S., and Sakaki, Y. (1985) Nucl. Acids Res. *13*, 7813–7827.
32. Scott, A. F., Schmeckpeper, B. J., Abdelrazik, M., Comey, C. T., O'Hara, B., Rossiter, J. P., Cooley, T., Heath, P., Smith, K. D., and Margolet, L. (1987) Genomics *1*, 113–125.
33. Bird, A. P. (1980) Nucl. Acids Res. *8*, 1499–1504.
34. Bird, A. P. (1986) Nature *321*, 209–213.
35. Gardiner-Garden, M., and Frommer, M. (1987) J. Mol. Biol. *196*, 261–282.
36. Rogers, J. H. (1985) International Review of Cytology *93*, 187–279.
37. Mizrokhi, L. J., Georgieva, S. G., and Ilyin, Y. V. (1988) Cell *54*, 685–691.
38. Jones, K. A., and Tjian, R. (1985) Nature *317*, 179–182.
39. Evans, T., DeChiara, T., and Efstratiadis, A. (1988) J. Mol. Biol. *199*, 61–81.
40. Mermod, N., Williams, T. J., and Tjian, R. (1988) Nature *332*, 557–561.
41. Fletcher, C., Heintz, N., and Roeder, R. G. (1987) Cell *51*, 773–781.
42. Toh, H., Hayashida, H., and Miyata, T. (1983) Nature 305, 827–829.
43. Toh, H., Kikuno, R., Hayashida, H., Miyata, T., Kugimiya, W., Inouye, S., Yuki, S., and Saigo, K. (1985) EMBO J. *4*, 1267–1272.
44. Skowronski, J., and Singer, M. F. (1986) Cold Spring Harb. Symp. quant. Biol. *51*, 457–464.
45. Furano, A. V., Robb, S. M., and Robb, F. T. (1988) Nucl. Acids Res. *16*, 9215–9231.