J Mol Evol (1987) 25:288–291

Journal of Molecular Evolution

Recent Insertion of an Alu Sequence in the Beta-Globin Gene Cluster of the Gorilla

G. Trabuchet, Y. Chebloune, P. Savatier, J. Lachuer, C. Faure, G. Verdier, and V.M. Nigon

Département de Biologie Générale et Appliquée, Université Claude Bernard—Lyon I, 43 Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France

Summary. We present the nucleotide sequence of a new Alu family member that lies between the delta- and beta-globin genes in gorilla DNA. The sequence exhibits 91% similarity with a consensus sequence of the Alu family. It is flanked by a perfect repetition of a 16-nucleotide target sequence and terminates with 24 adenylic residues. As this sequence is absent at this locus in other primate DNAs, its insertion occurred less than 8 million years ago, thus supporting the idea that Alu sequences are still mobile elements in the hominoid genome.

Key words: DNA sequencing — DNA evolution — Beta-globin gene — Alu repeats

Introduction

The Alu sequences constitute a family of small repetitive sequences interspersed throughout the primate genome. Their number has been estimated in the range of 300,000–900,000 copies in the human genome (Schmid and Jelinek 1982; Hwu et al. 1986), and evidence suggests that these repetitive sequences could be mobile and transpose through the genome. First, there is high variability among species in the number of Alu repeats (as estimated by titration of total DNA with an Alu-specific probe) in humans, chimpanzees, gorillas, and orangutans. This variability among closely related species suggests that many insertions or deletions of these sequences have occurred during the recent evolution of hominoids (Hwu et al. 1986). Second, the families of small interpersed repeats observed in the rabbit genome do not exhibit any homology with the Alu repeats (Hardison and Printz 1985), whereas significant homologies (in the range of 70–80%) are seen between intergenic sequences of rabbit and human at the beta-globin locus (Trabuchet et al., to be published).

In two instances, insertion of an Alu sequence into highly repetitive DNA sequences has been documented and suggests a recent insertion event. First, Grimaldi and Singer (1982) found an Alu repeat interrupting an alpha satellite DNA segment of the African green monkey and estimated the moment of this transposition event to be less than 12 million years (Myr) ago. Second, an M2 repeat, the type 2 Alu sequence in the mouse, occurs in the polymorphic repetitive sequence PR1 in the genome of the Balb/C mouse only. Its absence in other mouse strains demonstrates a recent integration event in the Balb/C mouse genome (Kominami et al. 1983). More recently, the transposition of an Alu sequence at the Mlvi-2 locus in a human B cell lymphoma has been documented, but in this third case it is not clear whether this modification represents a rare polymorphism in human DNA or a somatic rearrangement related to tumor induction (Economou-Pachnis and Tsichlis 1985).

Coupled with these indications of Alu mobility, there is also evidence of Alu immobility. For example, in two well-defined regions of the human and chimpanzee genomes, one containing the alpha cluster of globin genes and the other lying upstream from the delta-globin gene, there are seven Alu sequences. Neither of these regions has experienced additions, deletions, or concerted evolution of Alu sequences since the divergence of humans and chimpanzees (Maeda et al. 1983; Sawada et al. 1985).

Offprint requests to: G. Trabuchet

Fig. 1. **a** Map of the delta–beta-globin intergenic region in the gorilla. Striped boxes, exons of delta- or beta-globin genes; open boxes, introns; black box, Alu repeat; E, EcoRI; H, HindIII. **b** Strategy for sequencing the EcoRI–HindIII DNA fragment. The horizontal arrows below the map indicate sequence data obtained from individually generated Bal31 clones.

Much, therefore, remains to be learned about the tempo of Alu mobility.

Our paper describes the insertion of an Alu sequence between the delta- and beta-globin genes in the gorilla. The structure of this locus has been studied extensively and is conserved among higher primates (Barrie et al. 1981). Nucleotide sequences of this region previously have been determined from several primate DNAs: human (Poncz et al. 1983), chimpanzee (Savatier et al. 1985), and macaque (Savatier et al. 1987). These sequences are highly homologous and can be aligned without ambiguity by introducing only very small insertions or deletions. On the contrary, by restriction endonuclease mapping of genomic DNA, Zimmer et al. (1980) found a difference in the size of restriction fragments suggesting the presence of an extra piece of DNA at this locus in gorilla DNA. To analyze further this modification, we have cloned and sequenced this region and found an Alu repeat inserted at this locus. As this transposition occurred after the divergence of the gorilla from other apes, probably less than 8 Myr ago, we conclude that Alu repeats are still movable elements in hominoids.

Materials and Methods

A blood sample from a lowland gorilla (*Gorilla gorilla*) was provided by the Centre de Recherche Medicale de Franceville (Gabon). DNA preparation, blot-hybridization experiments, and cloning of the 6.0-kbp EcoRI DNA fragment containing the 5′ flanking region and the major part of the beta-globin gene (Fig. 1) have been already described (Savatier et al. 1987). The insert DNA was digested with the appropriate restriction endonuclease. Randomly terminated fragments were generated with Bal31 nuclease digestion followed by blunt-end ligation into the SmaI site of M13 mp10 or mp19 according to the method of Poncz et al. (1982). Sequencing was carried out by the method of Sanger et al. (1977) using ^{35}S-dATP essentially as described by Biggin et al. (1983). The strategy for sequencing is depicted in Fig. 1b.

Results

Genomic DNA was cut by EcoRI, electrophoresed, transferred onto nitrocellulose filter (Southern 1975), and hybridized with a specific beta-globin gene probe. A 6.0-kbp DNA fragment was observed in the gorilla sample, instead of the 5.5-kbp homologous fragment observed in human, chimpanzee, and macaque DNAs (Savatier et al. 1985, 1987). To analyze this size difference, we cloned the EcoRI fragment from gorilla DNA in the EcoRI site of the bacteriophage lambda gt WES lambda C. This fragment was then subcloned in the EcoRI site of the plasmid pEmbl 9 (Dente et al. 1983). A more detailed mapping of the plasmid DNA allowed us to locate the difference between human and gorilla in the 5′ part of the 6.0-kbp EcoRI DNA fragment in an EcoRI–HindIII DNA fragment (Fig. 1a). The size of the latter was found to be 1.9 kbp in gorilla and 1.6 kbp in human.

The nucleotide sequence (Fig. 2) reveals an inserted piece of DNA of 316 bp in the gorilla. Homology between the sequence of this insert and a consensus sequence of the Alu family was 91% (Schmid and Shen 1985). Moreover, the inserted Alu sequence is flanked on each side by the exact direct repetition of 16 nucleotides. Only one copy of the 16-nucleotide stretch is present in human DNA, which demonstrates that this target sequence has been duplicated during the insertion of the Alu repeat. Finally there is a perfect run of 24 adenylic residues at the 3′ end of the inserted sequence. All these features are characteristic of Alu family members of repetitive sequences (Schmid and Shen 1985).

The nucleotide sequence of gorilla DNA was also established on both sides of the insert and compared with the homologous human sequence over 1110 nucleotides. Nucleotide sequence divergence between the two species was calculated to be 2.0%, all

the differences consisting of substitutions or insertions/deletions of one nucleotide. Similar values in the range of 1.3–2.9% have been observed between these two species for putatively neutral DNA sequences (Savatier et al. 1987) or for unique genomic DNA sequences as estimated by delta-T50 measurements (Sibley and Ahlquist 1984; O'Brien et al. 1985).

Discussion

A new Alu repeat has been inserted in the intergenic sequence between the delta- and beta-globin genes in gorilla DNA. This insert is located 3.3 kbp upstream from the beta-globin gene. The structure of this region has been studied extensively, and the nucleotide sequence is known in human (Poncz et al. 1983), chimpanzee (Savatier et al. 1985), and macaque (Savatier et al. 1987). As the Alu sequence is absent at this locus in every primate species previously analyzed, we are confident that the insertion event occurred during evolution of the gorilla after the separation of this species from the other Hominidae.

The possibility exists that this Alu insert could be a polymorphism in gorilla. Indeed, Barrie et al. (1981) mapped this region in several apes and found no differences in the size of restriction fragments between gorilla and human DNA. Nevertheless, this last observation is at variance with those of Zimmer et al. (1980) and with ours, and could indicate the existence of a polymorphism in the gorilla. Whatever the case, the Alu sequence insertion is a recent event and is at most 8 Myr old, the estimated date for the gorilla lineage divergence (Sibley and Ahlquist 1984).

Two other characteristics of the new Alu sequence are in agreement with a recent insertion event. First, the gorilla Alu sequence is flanked by the perfect repetition of 16 nucleotides. The same nucleotide sequence is present only once in human DNA, and thus was duplicated during the integration event. Following the duplication, random mutation of these sequences could have occurred, so it has been proposed that nucleotide sequence divergence between the two flanking direct repeats be used as a measure of the date of insertion of the Alu repeat (Fukumaki et al. 1982; Schmid and Shen 1985). The fact that the two flanking sequences are perfectly identical in gorilla DNA argues for a recent insertion of the new Alu sequence. Second, this Alu repeat ends with an exact stretch of 24 adenylic residues. It is a common characteristic of Alu family sequences to terminate with an oligo-dA rich 3′ terminus. Insertion of other nucleotides at this place could result from mutations during aging of the Alu sequence.

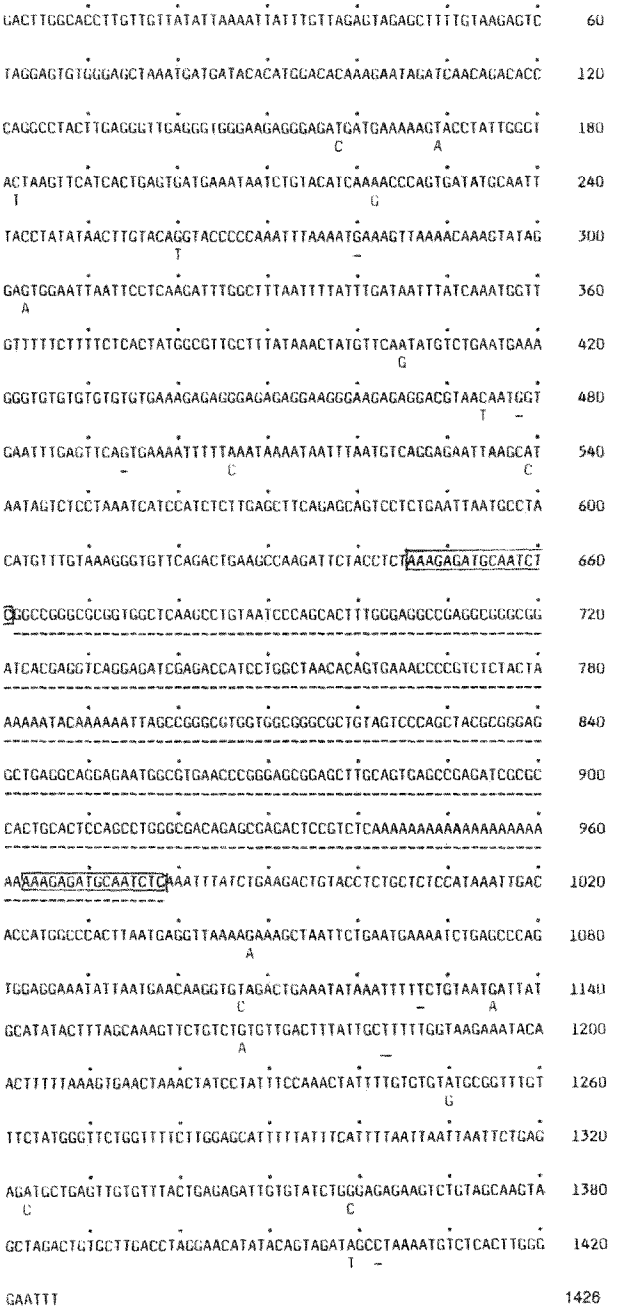

Fig. 2. Nucleotide sequence of a part of the EcoRI–HindIII gorilla DNA fragment including the Alu repeat. The gorilla sequence (upper line) is compared to the homologous human sequence (lower line). On the latter, only differences are indicated by the appropriate base and deletions by a dash. The direct repetitions flanking the Alu sequence are boxed.

As a consequence of its recent insertion, we believe that this Alu sequence is very similar to that from which it originated. Assuming it diverged at the same rate as the putatively neutral neighboring DNA, one can estimate that there is a maximum of three nucleotide differences between the present-day sequence and its progenitor. The calculation assumes that there is 2% nucleotide divergence between human and gorilla sequences, thus 1% be-

tween gorillas and their last common ancestor, which leads to three nucleotides different over 280 bp. In this respect the present sequence should be useful in studying the phylogeny of the Alu family.

Acknowledgments. We are grateful to Dr. Ivanoff (Centre de Recherches Médicales de Franceville, Gabon) for providing us with blood samples from the gorilla. We thank Dr. Y. Quentin for helpful discussions and the Institut d'Evolution Moléculaire and CITI2 for the use of their computer facilities. This work was supported by research grants from INSERM, CNRS, and CEE.

References

Barrie PA, Jeffreys AJ, Scott AF (1981) Evolution of the beta-globin gene cluster in man and in the primates. J Mol Biol 149:319–336

Biggin MD, Gibson TJ, Hong G (1983) Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc Natl Acad Sci USA 80:3963–3965

Dente L, Cesarini G, Cortese R (1983) pEmbl: a new family of single stranded plasmids. Nucleic Acids Res 11:1645–1655

Economou-Pachnis A, Tsichlis PN (1985) Insertion of an Alu Sine in the human homologue of the Mlvi-2 locus. Nucleic Acids Res 13:9379–9387

Fukumaki Y, Collins F, Kole R, Stoeckert CJ, Jagadeeswaran P, Duncan CH, Weissman SM (1982) Sequences of human repetitive DNA, non alpha-globin genes, and major histocompatibility locus genes: I. Repetitive DNA. Cold Spring Harbor Symp Quant Biol XLVII:1079–1081

Grimaldi G, Singer MF (1982) A monkey Alu sequence is flanked by 13-base pair direct repeats of an interrupted alpha-satellite DNA sequence. Proc Natl Acad Sci USA 79:1497–1500

Hardison RC, Printz R (1985) Variability within the rabbit C repeats and sequences shared with other SINES. Nucleic Acids Res 13:1073–1088

Hwu HR, Roberts JW, Davidson EH, Britten RJ (1986) Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. Proc Natl Acad Sci USA 83:3875–3879

Kominami R, Muramatsu M, Moriwaki K (1983) A mouse type 2 Alu sequence (M2) is mobile in the genome. Nature 301:87–89

Maeda N, Bliska JB, Smithies O (1983) Recombination and balanced chromosome polymorphism suggested by DNA sequences 5′ to the human delta-globin gene. Proc Natl Acad Sci USA 80:5012–5016

O'Brien SJ, Nash WG, Wildt DE, Bush ME, Benveniste RE (1985) A molecular solution to the riddle of the giant panda's phylogeny. Nature 317:140–144

Poncz M, Solowiejczyk D, Ballantine M, Schwartz E, Surrey S (1982) "Nonrandom" DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. Proc Natl Acad Sci USA 79:4298–4302

Poncz M, Schwartz E, Ballantine M, Surrey S (1983) Nucleotide sequence analysis of the delta–beta globin gene region in humans. J Biol Chem 258:11599–11609

Sanger F, Coulson AR, Barrell BG, Smith AJH, Roe BA (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Savatier P, Trabuchet G, Faure C, Cheboune Y, Gouy M, Verdier G, Nigon VM (1985) Evolution of the primate beta-globin gene region: high rate of variation in CpG dinucleotides and in short repeated sequences between man and chimpanzee. J Mol Biol 182:21–29

Savatier P, Trabuchet G, Cheboune Y, Faure C, Verdier G, Nigon VM (1987) Nucleotide sequence of the delta–beta globin intergenic segment in macaque: structure and evolutionary rate in higher primates. J Mol Evol 24:297–308

Sawada I, Willard C, Shen CKJ, Chapman B, Wilson AC, Schmid CW (1985) Evolution of Alu family repeats since the divergence of human and chimpanzee. J Mol Evol 22:316–322

Schmid CW, Jelinek WR (1982) The Alu family of dispersed repetitive sequences. Science 216:1065–1070

Schmid CW, Shen C-KJ (1985) The evolution of interspersed repetitive DNA sequences in mammals and other vertebrates. In: Macintyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 323–358

Sibley CG, Ahlquist JE (1984) The phylogeny of the hominoid primates as indicated by DNA–DNA hybridization. J Mol Evol 20:2–15

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC (1980) Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. Proc Natl Acad Sci USA 77: 2158–2162

Received January 6, 1987/Accepted January 16, 1987