doi:10.1016/j.jmb.2005.03.058 J. Mol. Biol. (2005) 350, 200–214

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

An Unusual Primate Locus that Attracted Two Independent Alu Insertions and Facilitates their Transcription

A. Ludwig, T. S. Rozhdestvensky, V. Yu. Kuryshev, J. Schmitz and J. Brosius*

Institute of Experimental Pathology, ZMBE, University of Münster, Von-Esmarch-Str. 56 D-48149 Münster, Germany

*Corresponding author

BC200 RNA, a neuronal, small non-messenger RNA that originated from a monomeric Alu element is specific to anthropoid primates. Tarsiers lack an insert at the orthologous genomic position, whereas strepsirrhines (Lemuriformes and Lorisiformes) acquired a dimeric Alu element, independently from anthropoids. In *Galago moholi*, the CpG dinucleotides are conspicuously conserved, while in *Eulemur coronatus* a large proportion is changed, indicating that the *G. moholi* Alu is under purifying selection and might be transcribed. Indeed, Northern blot analysis of total brain RNA from *G. moholi* with a specific probe revealed a prominent signal. In contrast, a corresponding signal was absent from brain RNA from *E. coronatus*. Isolation and sequence analysis of additional strepsirrhine loci confirmed the differential sequence conservation including CpG patterns of the orthologous dimeric Alu elements in Lorisiformes and Lemuriformes. Interestingly, all examined Alu elements from Lorisiformes were transcribed, while all from Lemuriformes were silent when transiently transfected into HeLa cells. Upstream sequences, especially those between the transcriptional start site and −22 upstream, were important for basal transcriptional activity. Thus, the BC200 RNA gene locus attracted two independent Alu insertions during its evolutionary history and provided upstream promoter elements required for their transcription.

Keywords: small non-messenger RNA; Alu elements; primate evolution; retronuons; neuronal transcription

Introduction

The process of retroposition (reverse transcription of RNA into cDNA and concomitant genomic integration)[1–3] has contributed as much as 90% of DNA sequences in mammalian genomes.[4] Rather than mere genomic "junk", retronuons[5,6] are often recruited or exapted into novel functions as they are mobile carriers of potential control sequences (sometimes after a few minor alterations) and, thus, can equip existing genes with novel regulatory elements.[5–15] Retronuons also provide existing genes with novel protein-coding domains,[14,16,17] and even generate novel genes,[5,18,19] including those encoding non-messenger RNAs.[20,21]

The most abundant class of retronuons in vertebrate genomes are the short interspersed elements (SINEs).[22] Most SINEs originated by retroposition of tRNAs,[23] tRNA-related RNAs or signal recognition particle (SRP) RNA (7SL RNA).[24–27] The SRP RNA-related Alu elements constitute the major SINE family in primates and exist in both monomeric and dimeric form.[28,29] Only a limited number of active Alu source (master)

Present addresses: A. Ludwig, Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland; V. Yu. Kuryshev, Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany.

Abbreviations used: SINE, short interspersed element; SRP, signal recognition particle; snoRNA, small nucleolar RNA; fl, full-length; sc, small cytoplasmic; DSE, distal sequence element; FLAM, free left Alu monomer; PSE, proximal sequence element.

E-mail address of the corresponding author: rna.world@uni-muenster.de

genes have given rise to the $>10^6$ Alu copies in primate genomes by successive waves of retropositional activity.[26,27,30] In certain lineages and at certain times, multiple source genes can be active in parallel, as shown for the BC1 RNA-related retronuons (ID elements) in rodents[8,9] and Alu elements in primates.[8,9,31] Apart from selective constraint on the source RNA, the redundancy of active elements might contribute to survival of retronuons in genomes. Despite the huge number of Alu element copies in primate genomes, only a few Alu transcripts have been observed *in vivo*.[28,32–34] Clearly, 5′ flanking sequences are important for autonomous transcription of SINEs.[21,34–39] Therefore, the genomic environment, which is unique for each individual SINE, plays a decisive role in whether a SINE is transcriptionally active or silent.[6,8,40]

BC200 RNA, a neuronal, small non-messenger RNA, originated from a free left Alu monomer (FLAM)-C-like monomeric Alu element.[21,41] The BC200 RNA gene represents one of the rare members of this repetitive DNA family that is transcriptionally and retropositionally active.[42–44] Like its possible functional analogue in rodents, BC1 RNA,[20] BC200 RNA is transcribed in neurons[45] predominantly by RNA polymerase III[21] and transported to dendrites as a ribonucleoprotein particle.[46–49]

The gene encoding BC200 RNA has been mapped to chromosome 2p16,[50] but the human genome sequence places it to band 2p21 between the CALM2 (calmodulin 2) and the TACSTD1 (tumour-associated calcium signal transducer 1) genes.[51] The gene is conserved in all anthropoids including Old World monkeys, New World monkeys and hominoids.[43] V. Yu. K. *et al*. showed that the BC200 RNA gene is absent from orthologous loci in representatives of the strepsirrhines (Lorisiformes and Lemuriformes) and tarsiers, suggesting that the gene has arisen on the lineage leading to anthropoids after the separation of Strepsirrhini and Haplorrhini (tarsiers and anthropoids; for phylogenetic affiliations see Schmitz *et al*.[52]).[44] Surprisingly, the *G. moholi* (lorisiform) and *E. coronatus* (lemuriform) loci harbor a dimeric Alu element at the orthologous position. This sets apart the aforementioned strepsirrhines from tarsiers, whose orthologous loci are devoid of any repetitive element.[44] Interestingly, the *G. moholi* Alu (referred to as G22) features 21 preserved CpG doublets (usually the first to change in genomic sequences released from purifying selection),[53,54] whereas the *E. coronatus* Alu (referred to as L13) features only 11 CpG doublets. The conservation of CpG doublets is an indication for transcriptional activity and functionality, as had been shown for the BC200 RNA gene.[41,43] This prompted us to examine whether G22 is transcribed in the *G. moholi* brain, akin to BC200 RNA in anthropoids. In order to examine the strepsirrhine locus corresponding to the BC200 RNA gene locus in more detail, we PCR-amplified and sequenced the locus of five other Lemuriformes (including *Daubentonia madagascariensis*), three other Lorisiformes, one additional tarsier species as well as the colugo, a non-primate, but presumably primate-related mammal (*Cynocephalus variegatus*, Dermoptera).[55]

Results

The *G. moholi* G22 Alu element is transcribed in the brain

The BC200 RNA gene is conserved and transcribed in the anthropoid brain,[43] but is absent from orthologous loci of *G. moholi*, *E. coronatus* and *T. syrichta*.[44] While the strepsirrhine (*G. moholi* and *E. coronatus*) loci harbour a dimeric Alu element at exactly the same chromosomal location, the *T. syrichta* locus is devoid of any repetitive element. Unlike the *E. coronatus* Alu L13, the *G. moholi* Alu G22 preserved its internal box A and B promoter elements and its CpG doublets. Thus, we first examined whether the G22 Alu would be transcriptionally active in the *G. moholi* brain, akin to BC200 RNA in anthropoids.

We performed Northern hybridization of total RNA isolated from various tissues of *G. moholi*, *E. coronatus*, *T. syrichta* and human, and probed for the G22 and L13 Alu elements using several complementary oligonucleotides (Figure 1). An oligonucleotide specific for the lorisiform G22 Alu element (G22_3′) detected a single band in the galago brain exhibiting an apparent size of ~340 nt (Figure 1(a)). A "generic" probe complementary to the Alu 5′ domain (Alu_5′) recognized, in addition to the G22 Alu RNA, a transcript with an apparent size of 300 nt. This transcript was detected in *E. coronatus* and *G. moholi* tissues, but not in *T. syrichta* and human samples (Figure 1(b)). The lack of any signal in tarsier and human tissues reduces the possibility of cross-hybridization to 7SL RNA (see Figure 3) and rather suggests that there is an additional distinct Alu retronuon expressed in the *G. moholi* and *E. coronatus* tissues examined.

In order to probe for the *E. coronatus* L13 Alu element, we used two distinct oligonucleotides (L13_insert and L13_3′). Both oligonucleotides failed to detect a transcript in *E. coronatus* tissues that might correspond to the L13 Alu (data not shown). Unexpectedly, however, the oligonucleotide probe L13_insert detected a band with an apparent size of 210 nt in human brain and, to a lesser degree, in *G. moholi* brain and liver (Figure 1(c)). This points to yet another distinct transcript that is not necessarily Alu-related. With respect to the *G. moholi* G22 Alu element, our data indicate that it is prevalently, perhaps exclusively, transcribed in the brain, while the *E. coronatus* Alu L13 is transcriptionally silent. This observation prompted us to examine whether other lorisiform and lemuriform species also harbour a dimeric Alu element at that chromosomal position. Thus, we PCR-amplified, sequenced and analysed

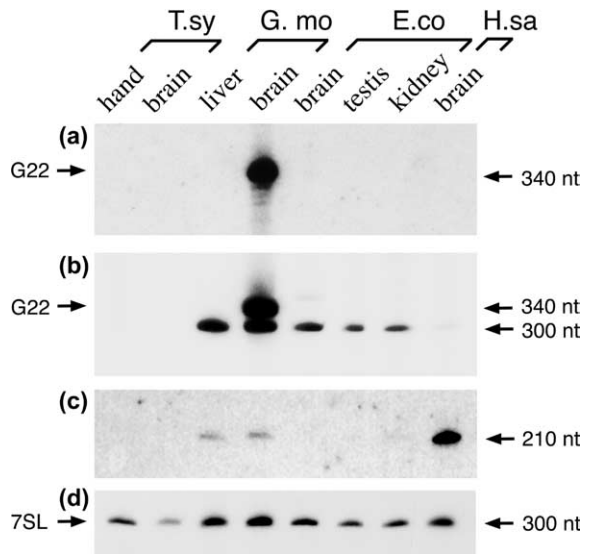


Figure 1. The *G. moholi* Alu element G22 is transcribed in the galago brain. Total RNA isolated from various tissues of *T. syrichta* (T.sy), *G. moholi* (G.mo), *E. coronatus* (E.co) and human (H.sa) was transferred to nylon membrane and probed with various oligonucleotides. (a) The G22-specific oligonucleotide G22_3′ recognizes a single band with an apparent size of 340 nt in *G. moholi* brain, but not in liver. (b) The oligonucleotide Alu_5′ (a "generic" Alu probe) detects a band with an apparent size of 340 nt in *G. moholi* brain and a transcript of 300 nt in *G. moholi* liver and brain, and in all *E. coronatus* tissues examined. (c) The L13-specific oligonucleotide L13_insert (a sequence unrelated to Alu consensus sequences) fortuitously recognizes a transcript of 210 nt in *G. moholi* liver and brain, and in human brain, but does not detect any transcript in *E. coronatus* tissues. (d) The 7SL RNA oligonucleotide 7SL_3′ was used as loading control.

orthologous BC200 loci from additional strepsirrhine species.

Orthologous loci of the anthropoid BC200 RNA gene locus in strepsirrhines harbour an AluJo-like element

We PCR amplified (see Supplementary Figure 1) and sequenced loci orthologous to the anthropoid BC200/galago G22 gene locus from three other Lorisiformes (*Loris tardigradus, Nycticebus pygmaeus, Otolemur garnettii*), five Lemuriformes (*Cheirogaleus medius, Lepilemur dorsalis, Lepilemur ruficaudatus, Propithecus verreauxi, Daubentonia madagascariensis*), one additional tarsier (*Tarsius bancanus*) and the non-primate colugo (*Cynocephalus variegatus*, Dermoptera). All lorisiform and lemuriform species examined contain a dimeric Alu element at the orthologous position. In contrast, the *T. bancanus* and the colugo loci are devoid of any insertion. For further insights into the BC200 RNA gene locus, we aligned the strepsirrhine loci with the previously published anthropoid BC200 loci,[43] the tarsier loci as well as the outgroup sequences of colugo and mouse (Figure 2).

As the sequences of the dimeric Alu elements in the two strepsirrhine lineages are quite divergent, we asked whether the lorisiform G22 and lemuriform L13 Alu elements descend from a common Alu element or from independent progenitors. To this end, we aligned and compared the G22 and L13 Alu elements with the human AluJo consensus sequence, which is believed to be one of the most ancient dimeric Alu elements,[29] and thus the closest relative of the strepsirrhine Alu elements examined here,[44] as well as with lorisiform and lemuriform Alu consensus sequences (GS1, Gal and Lem-S; Figure 3). Inclusion of these consensus sequences aided in finding shared diagnostic positions (single nucleotide substitutions, insertions or deletions, when compared to the consensus)[26,41] between the lorisiform G22 and lemuriform L13 Alu elements. This sequence analysis revealed at least six common diagnostic positions for the G22 and L13 Alu elements, which are not shared with any of the consensus sequences, including the AluJo sequence. We found at least ten additional diagnostic positions that the G22 and L13 Alu elements share with the lorisiform and lemuriform consensus sequences (Gal, GS1 and Lem-S), but not with the AluJo consensus. Furthermore, G22 and L13 Alu elements have other hallmarks in common, such as a 2 bp deletion at position 114, which is also present in the lorisiform and lemuriform consensus sequences, and a CCC trinucleotide adjacent to the A-rich region that links the two monomer domains (~ position 220, Figure 3). This analysis suggests that the lorisiform G22 and lemuriform L13 Alu elements descended from a common ancestral AluJo-like element that integrated prior to strepsirrhine divergence.

The G22 and L13 Alu elements are not equally conserved. Sequence comparison using uncorrected pairwise distances (p) as implemented in PAUP revealed that, even after excluding the distant aye-aye, the five lemuriform Alu sequences are about threefold more divergent than the four lorisiform Alu elements (data not shown). Furthermore, while the G22 Alu elements preserved the internal A and B-box promoter elements, the L13 Alu elements display various nucleotide substitutions and deletions (Figure 3). MFOLD-analysis[56] further illustrated the conservation of a prototypical Alu secondary structure in the G22 Alu elements but not in the L13 Alu elements (data not shown). Importantly, the two Alu families exhibit pronounced differences in the preservation of their respective CpG doublets. The Galagidae *G. moholi* and *O. garnettii* preserved 21 and 20 CpG doublets, respectively, while the Lorisidae *N. pygmaeus* and *L. tardigradus* preserved 20 and 16 CpG dinucleotides, respectively. In striking contrast, the Lemuriformes *E. coronatus*, *P. verreauxi*, *C. medius*, *L. ruficaudatus* and *L. dorsalis* preserved only 11, 9, 11, 11 and 10 CpG doublets, respectively. Surprisingly, the *D. madagascariensis* Alu element

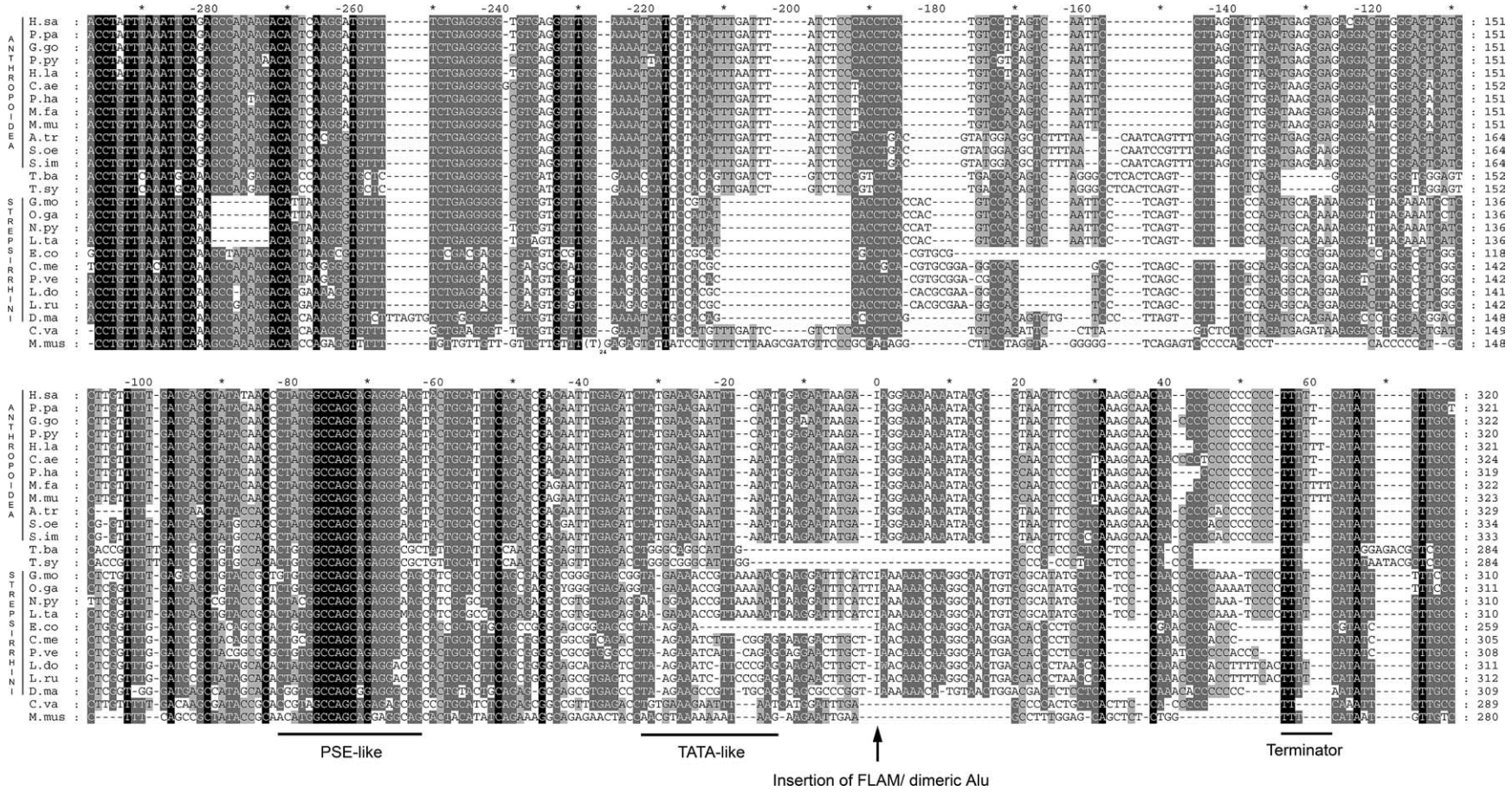


Figure 2. Nucleotide sequence alignment of the central core sequence surrounding the anthropoid BC200 RNA gene locus. Orthologous flanking sequences of the BC200 RNA gene locus available in strepsirrhines (Lemuriformes and Lorisiformes), tarsiers, New and Old World monkeys as well as hominoids including human[43] were aligned (see Materials and Methods) together with orthologous loci of the non-primate outgroups *Cynocephalus variegatus* (colugo) and *Mus musculus*. The integration point (I) of the FLAM in anthropoids (resulting in the BC200 gene) and the dimeric AluJo-like element in strepsirrhines (giving rise to the G22 and L13 Alu elements) is depicted. Putative regulatory elements (PSE, TATA-like element and the terminator sequence) are denoted. Sequences identical in all species are indicated in black, positions with $\geq$80% and $\geq$60% similarity are shaded in dark grey and light grey, respectively. Note the deletion around the integration site in tarsiers and the preservation of the TATA-like sequence in anthropoids and Lorisiformes. For abbreviations, see Materials and Methods.

features 24 CpG dinucleotides (nine of which, however, are not shared by any of the Alu consensus sequences). The combined information strongly indicates that the lorisiform G22 Alu elements have been under purifying selection and may, thus, be transcriptionally and functionally active.

G22 but not L13 Alu elements are transcribed in HeLa cells

We sought to determine whether transcriptional activity is restricted to the lorisiform G22 Alu elements or *vice versa*, whether all lemuriform Alu elements examined are inactive. In order to circumvent the need for tissue samples from all these animals, we performed transcriptional analysis by transient transfections of individual Alu elements including sufficient flanking sequences (~270 bp upstream and ~150 bp downstream) into HeLa cells. The *T. bancanus* construct (lacking the Alu element) was used as a negative control in transfections and a plasmid construct containing the mouse small nucleolar RNA (snoRNA) MBII-52 under control of the CMV promoter[57] was used as a positive control and was always co-transfected with the strepsirrhine loci.

Transient transfection into HeLa cells demonstrated clearly that all lorisiform G22 Alu elements were effectively transcribed. Somewhat unexpectedly, the G22 Alu elements produced transcripts of different sizes (Figure 4). As expected, the *G. moholi* and *L. tardigradus* Alu elements exhibited an apparent size of 340 nt. In contrast, those of *O. garnettii* and *N. pygmaeus* generated transcripts with apparent sizes of 360 nt and 380 nt, respectively. From the sequence information and location of RNA polymerase III termination sites (a stretch of at least four consecutive thymidine residues, Figure 2), all transcripts should be in a similar size range. Although we are unable to explain this discrepancy, one explanation could be post-transcriptional addition of nucleotides.[58,59]

In contrast to the lorisiform G22 Alu elements, none of the lemuriform L13 Alu elements, including the *Daubentonia* Alu, was transcriptionally active in HeLa cells (Figure 5). Neither an oligonucleotide probe common to Alu elements (Alu_5′, Figure 5(a)), nor one complementary to the *E. coronatus* insert (L13_insert, Figure 5(b)) or the right arm of the lemuriform Alu elements (L13_3′, data not shown) produced a signal corresponding to the lemuriform Alu elements. Moreover, a chimeric construct, in which the *G. moholi* 5′ flanking sequence was fused to the *E. coronatus* Alu element (including its 3′ flanking sequence), was transcriptionally inactive (lane GF/LA, Figure 5), as were similar chimeric constructs between the *G. moholi* 5′ flank and the *P. verreauxi* and the *D. madagascariensis* L13 Alu elements, respectively (data not shown). These data, together with the results shown in Figure 1, show that transcriptional activity is restricted to the lorisiform G22 Alu elements.

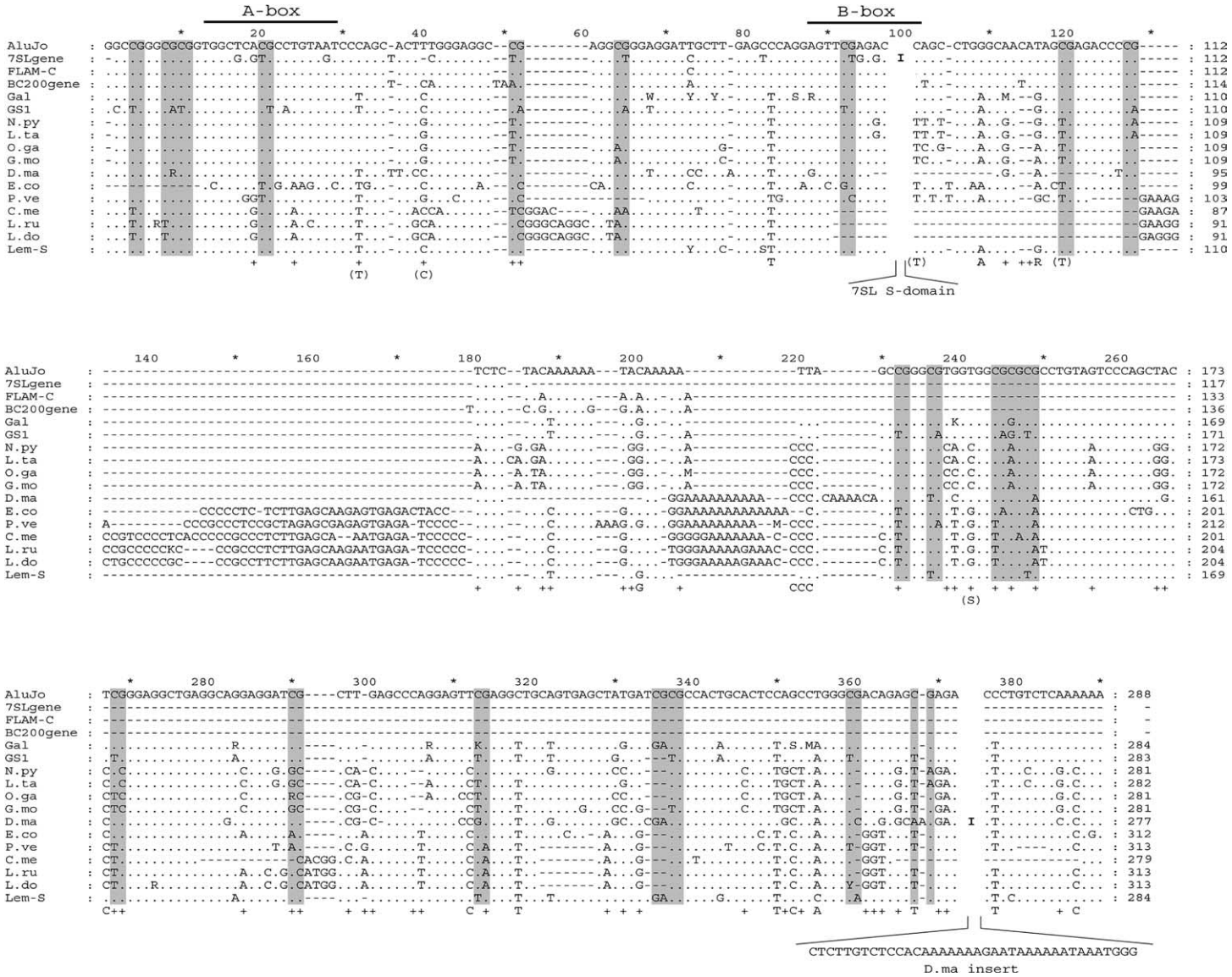
We detected transcripts with apparent sizes of 300 nt and 110 nt in transfected and mock-transfected HeLa cells (Figure 5(a)) using the Alu_5′ oligonucleotide (a generic Alu probe complementary to the left arm). We suggest that these hybridization signals correspond to full-length (fl) and small cytoplasmic (sc) Alu transcripts endogenous to HeLa cells. Transfection of the *G. moholi* Alu element produced, in addition to the 340 nt long G22 flAlu transcript, also a band of approximately 110 nt, which co-migrates with the aforementioned scAlu from HeLa cells (Figure 5(a)). The same pattern was observed in *G. moholi* brain (data not shown). This suggests that the G22 Alu element is, as human transcribed Alu elements, subject to post-transcriptional processing,[28] which is supported by the detection of possible processing intermediates (Figures 1, 4–6).

With oligonucleotide L13_insert as probe, two fortuitous transcripts with apparent sizes of 210 nt and 130 nt were detected in transfected as well as mock-transfected HeLa cells (Figure 5(b)). The 210 nt transcript was detected also in total RNA isolated from *G. moholi* tissues and in human brain (see Figure 1). Clarification of the nature of these RNAs requires further experimentation.

Transcription of the *G. moholi* G22 Alu element requires 5′ flanking sequences

After establishing that the lorisiform G22 Alu elements are expressed, we carried out experiments to narrow down the promoter elements necessary for transcription and used the *G. moholi* locus as a model. As 5′ flanking sequences were shown to augment transcription of the SRP RNA gene,[60] Alu elements[21,34,35,37] and SINE master genes such as the rodent BC1 RNA gene,[36] we initially screened the lorisiform and anthropoid 5′ flanking sequences for the presence of putative promoter elements. Upstream sequence elements preserved in Lorisiformes and Anthropoidea show similarity to known external promoter elements found in RNA polymerase III and II genes (Figure 2). The most striking feature is that lorisiforms and anthropoids harbour a conserved TA-rich region between positions −20 and −30, akin to TATA motifs. Second, a conserved sequence block reminiscent of a proximal sequence element (PSE)[61,62] is situated about 70 bp upstream of the insertion point. Third, 220 bp upstream of the integration point, we found a distal sequence element (DSE)-like motif, which harbours an imperfect octamer motif (ATTTAAAT in hominoids and GTTTAAAT in Lorisiformes, New and Old World monkeys) and a sequence motif (GGGG(T/C)GTG(AG/GT)GGTTGGAAA AT) reminiscent of a Staf-responsive element.[63]

To test whether the conserved regions present in the *G. moholi* 5′ flanking sequences influence transcription of the G22 Alu element, we designed several plasmid constructs, in which the 5′ flanking


A-box
B-box
7SL S-domain
D.ma insert
CTCTTGTCTCCACAAAAAAAGAATAAAAAATAAATGGG
AluJo
7SLgene
FLAM-C
BC200gene
Gal
GS1
N.py
L.ta
O.ga
G.mo
D.ma
E.co
P.ve
C.me
L.ru
L.do
Lem-S

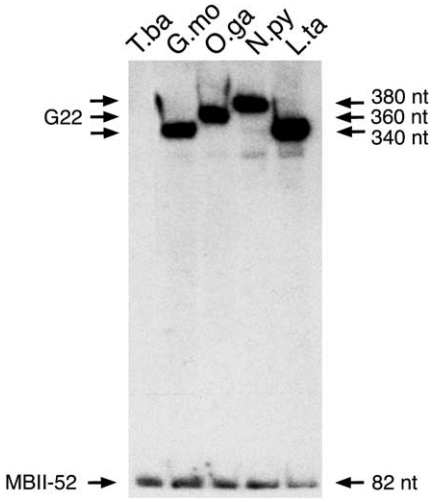


Figure 4. Lorisiform G22 Alu elements are transcribed in HeLa cells. Northern blot hybridization of total RNA derived from the transient transfection of lorisiform G22 Alu elements in HeLa cells. From left to right: *T. bancanus* (negative control), *G. moholi* G22, *O. garnettii* G22, *N. pygmaeus* G22, *L. tardigradus* G22. The membrane was probed with the oligonucleotides G22_3′ and MBII-52_P19. Transcripts of the G22 Alu elements and the co-transfected MBII-52 snoRNA are denoted by arrows.

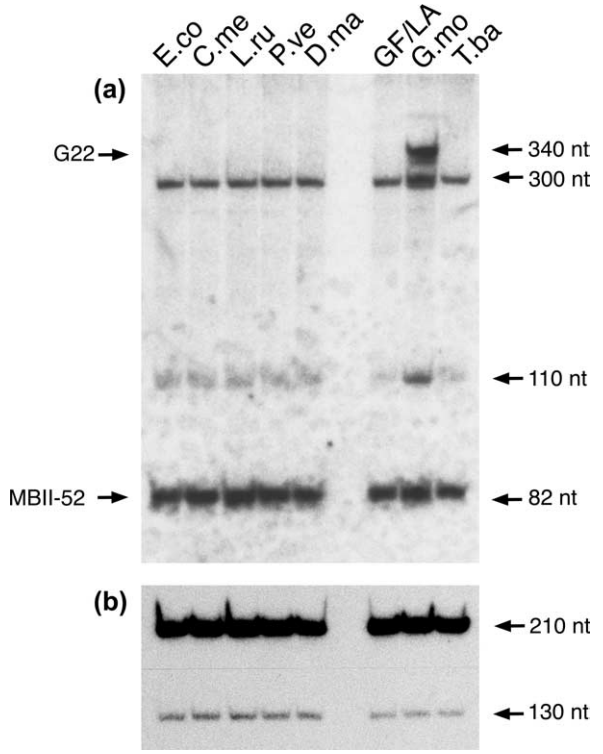


Figure 5. Lemuriform L13 Alu elements are not transcribed in HeLa cells. Northern blot hybridization of total RNA derived from the transient transfection of lemuriform L13 Alu elements in HeLa cells. From left to right: *E. coronatus* L13, *C. medius* L13, *L. ruficaudatus* L13, *P. verreauxi* L13, *D. madagascariensis* L13, GF/LA chimeric construct (the *G. moholi* flank fused to the *E. coronatus* L13 Alu element), *G. moholi* G22 (positive control), *T. bancanus* (negative control). (a) Membrane probed with oligonucleotides Alu_5′ and MBII-52_P19. The Alu_5′ oligonucleotide recognized a transcript with an apparent size of 340 nt only in the *G. moholi* transfection and a 300 nt transcript in all transfections (see also Figure 1). In addition, a smaller transcript with an apparent size of 110 nt was detected, whose intensity was stronger in the *G. moholi* transfection. The co-transfected MBII-52 snoRNA is denoted by an arrow. (b) The same membrane re-probed with the oligonucleotide L13_insert (specific for the *E. coronatus* insert). Two transcripts with apparent sizes of about 210 nt and 130 nt were recognized by this oligonucleotide, but no signal corresponding to the lemuriform L13 Alu elements could be detected.

sequence was gradually removed (Figure 6(c)). The DSE-like motif seemed negligible for transcription in this cell culture system, as a construct in which the 5′ flank terminated just upstream of the PSE (construct −86) was almost equally efficient at transcription as constructs bearing additional 5′ flanking sequences (constructs −207, −242, −270 and −383, Figure 6(a)). However, constructs lacking the PSE (−61, −41) showed reduced transcription levels and the deletion construct −22, which has all but the TATA-like sequence removed, showed a significant reduction in transcriptional activity when compared to the deletion mutant −86 (Figure 6(b) and (d)). Furthermore, removal of the TATA-like sequence resulted in a complete loss of transcriptional activity (construct +1). These data suggest that 5′ flanking sequences residing within the first 86 bp upstream of the putative transcription start site, and especially the TATA-like

Figure 3. Nucleotide sequence alignment of the strepsirrhine dimeric Alu elements G22 and L13 with the AluJo consensus sequence. Included are the human BC200 RNA, FLAM-C and 7SL RNA genes as well as consensus sequences for strepsirrhine Alu elements, GS1 (the consensus sequence for the Galagidae type I Alu family), Gal (a further consensus sequence for the Galagidae family) and Lem-S, a consensus sequence for a lemur Alu family. CpG dinucleotides conserved in the AluJo sequence are shaded in grey. RNA polymerase III promoter elements, A and B-boxes, are highlighted. Diagnostic nucleotide positions for either the lemuriform L13 or the lorisiform G22 Alu elements examined are indicated with +. Shared diagnostic positions between the L13 and G22 Alu elements, when compared to the consensus sequences, are denoted with the respective consensus nucleotide (A, T, G, or C). Polymorphic nucleotide positions are depicted using the following nomenclature: R = A or G; Y = C or T; S = G or C; W = A or T; K = G or T; M = A or C; N = any. The central region of the 7SL gene (corresponding to the S-domain in 7SL RNA) was omitted from the alignment (see position 99) and the duplication specific to *D. madagascariensis* was placed below the sequence (see position 373). For abbreviations, see Materials and Methods.

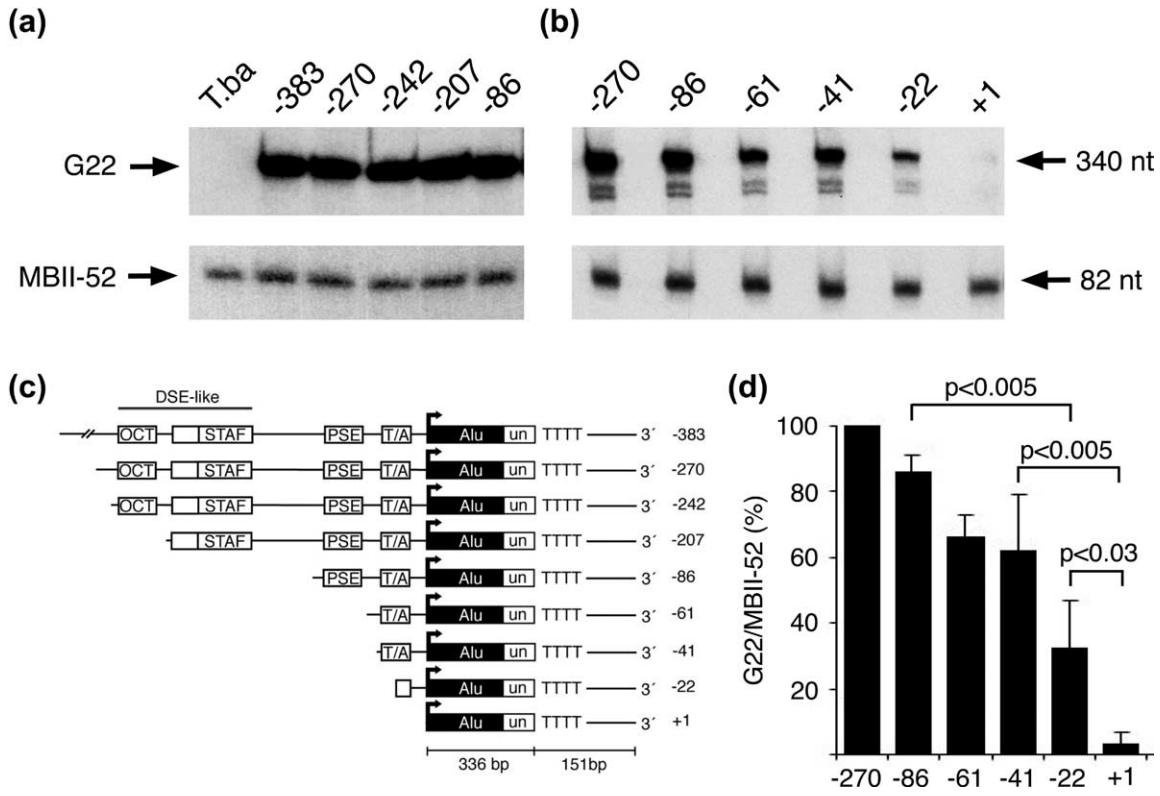


Figure 6. The *G. moholi* G22 Alu element requires 5′ flanking sequences for full transcriptional activity in HeLa cells. HeLa cells were transiently transfected with equal amounts of *G. moholi* G22 5′ deletion constructs and transcriptional activity was assessed by Northern blot analysis of total RNA. (a) Removal of the more remote 5′ sequence elements (see (c)) has no effect on G22 transcription. (b) Transcriptional activity of the G22 Alu decreases gradually upon removal of the 5′ flanking sequences that are located closer to the transcriptional start site (see also (c)). Membranes in (a) and (b) were probed with the oligonucleotides G22_3′ and MBII-52_P19. (c) A sketch showing putative promoter elements that were gradually omitted from the *G. moholi* 5′ flanking sequence in (a) and (b). (d) Relative transcriptional activities of the G22 5′ deletion constructs shown in (b) determined from three independent experiments.

sequence, are important for transcription of the G22 Alu in this cell culture system.

We then addressed the question of whether the tarsier and the colugo loci (two orthologous loci that have no Alu insertion) would support transcription of the *G. moholi* G22 Alu. To this end, we artificially inserted the *G. moholi* G22 Alu including its terminator sequence into the Alu-deficient *T. bancanus* and *C. variegatus* loci, respectively. We also wished to determine whether a given lemuriform flank can substitute for the *G. moholi* wild-type flank. For this purpose, we generated chimeric constructs that had either the *E. coronatus* or the *P. verreauxi* 5′ flank fused to the *G. moholi* G22 Alu. These chimeric constructs (depicted in Figure 7(a)) were transfected into HeLa cells and their respective transcriptional activities compared to the *G. moholi* constructs −86, −22 and +1 (Figure 7(b)). The *P. verreauxi* 5′ flank significantly enhanced transcription of the G22 Alu when compared to the level of the *G. moholi* +1 construct, whereas the *E. coronatus* flank, which lacks the TATA-like sequence, did not support transcription. Interestingly, the *G. moholi* G22 Alu was transcribed when being inserted into the colugo locus, although at lower levels than the *G. moholi* −22 construct. In contrast, the G22 Alu inserted into the tarsier locus reached transcription levels just above background. These data demonstrate that orthologous loci of non-lorisiform species have fortuitously maintained the potential to support transcription of an artificially inserted Alu retronuon.

Discussion

BC200 RNA is transcribed almost exclusively in the anthropoid brain[43] and arose *via* integration of a FLAM-C-like monomeric Alu element, probably after the tarsier divergence.[44] Here, we show that in the strepsirrhine lineage, prior to the divergence of Lorisiformes and Lemuriformes, the orthologous strepsirrhine locus had been subject to an independent insertion event of a dimeric Alu element, giving rise to the lorisiform G22 and lemuriform L13 Alu elements. Interestingly, while the G22 Alu elements are transcribed, the L13 Alu elements are transcriptionally inactive (summarized in Figure 8). Using a HeLa cell culture system, we demonstrate that certain upstream promoter elements are important for transcription of the *G. moholi* G22 Alu. Furthermore, orthologous loci of related

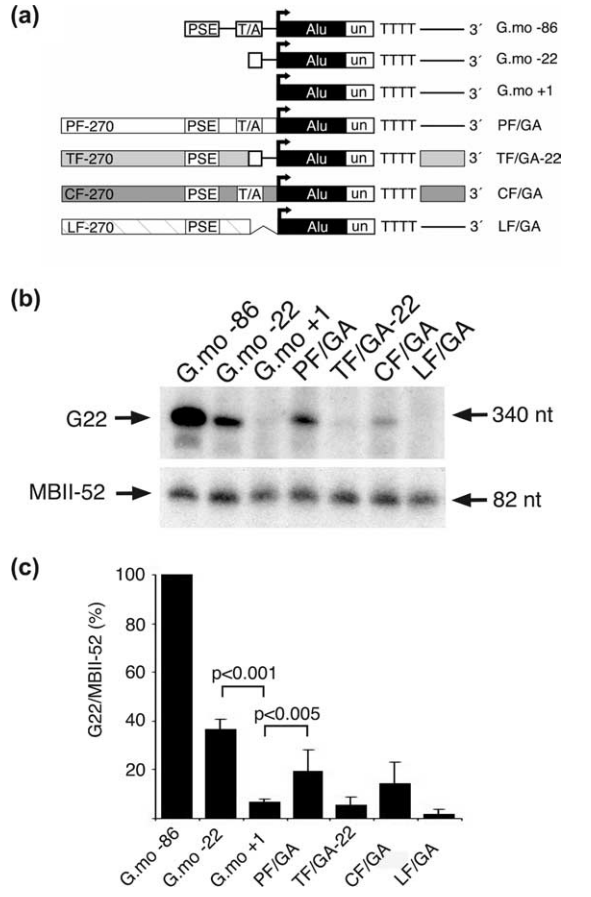


Figure 7. Orthologous 5′ flanking sequences from related species can support transcription of the *G. moholi* G22 Alu element in HeLa cells. (a) A sketch showing the *G. moholi* 5′ deletion constructs −86, −22 and +1, and the design of the chimeric constructs fusing the 5′ flanking sequences of *P. verreauxi* (PF/GA), *T. bancanus* (TF/GA-22), *C. variegatus* (CF/GA) and *E. coronatus* (LF/GA) to the *G. moholi* G22 Alu element. Note that the *E. coronatus* 5′ flanking sequence harbours a deletion of the T/A rich region (angled lines) and that the *G. moholi* T/A rich region was included in the TF/GA-22 construct (open box). (b) HeLa cells were transiently transfected with equal amounts of *G. moholi* G22 5′ deletion constructs and the chimeric constructs depicted in (a). Transcriptional activity was assessed by Northern blot analysis of total RNA. The membrane was probed with the oligonucleotides G22_3′ and MBII-52_P19. (c) Relative transcriptional activities of the chimeric constructs shown in (b) determined from three independent experiments.

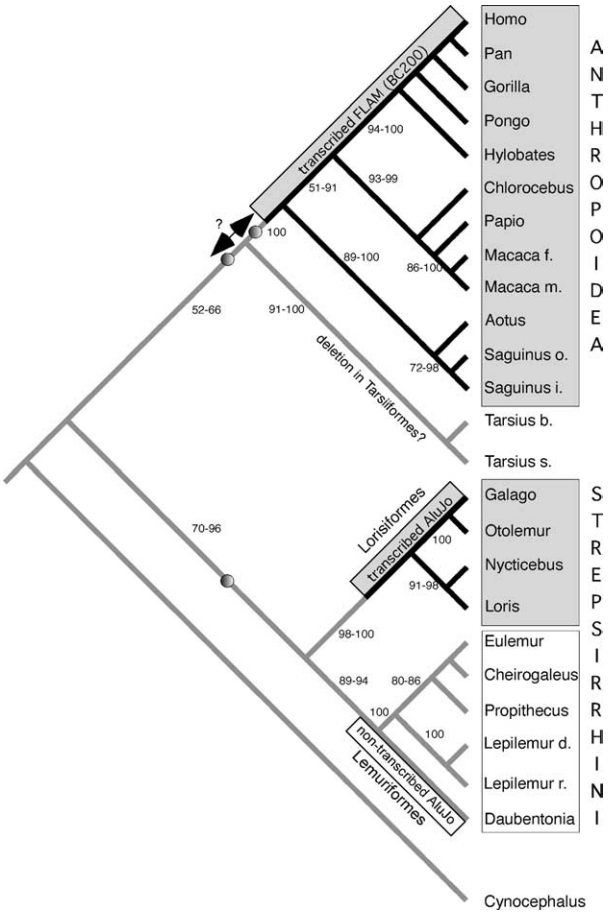


Figure 8. Phylogenetic reconstruction of the BC200 RNA gene locus. Orthologous flanking sequences of the BC200 RNA gene locus in primates were analysed. *C. variegatus* (colugo) was used as outgroup. The range of bootstrap or puzzling support values derived by the MP, NJ and ML method (see Materials and Methods) are given at the corresponding branches. The integration sites of FLAM (BC200) at the lineage leading to anthropoids and the AluJo in strepsirrhines are represented as dots. Because of the random deletion around the BC200 integration site in tarsiers the integration could alternatively have taken place before the tarsiers diverged. Branches with actively transcribed BC200 or Alu elements are indicated by bold lines.

species have residual potential to support transcription of an artificially inserted G22 Alu element.

A primate locus that attracted two independent Alu insertions

Alu elements preferentially insert into TT/AAAA or similar motifs surrounded by an otherwise GC-rich sequence.[64–66] It has been speculated that the purine–pyrimidine boundary may cause a kink in the DNA, which could attract proteins required for integration. Indeed, the TT/AAAA target is nicked by the L1 endonuclease.[67,68] The anthropoid and lorisiform loci as well as the colugo locus maintained remnants of the TT/AAAA motif at around −20 (Figure 2). This particular sequence landscape may have been advantageous for independent retronuon insertions at this locus.

The present data suggest that the BC200 RNA gene locus was targeted at least twice by related yet different retronuons during its evolutionary history, a dimeric Alu in strepsirrhines and a monomeric Alu element in anthropoids. Theoretically, an alternative scenario exists: It is known that existing

Alu elements can be replaced by gene conversion with other Alu elements.[21,69] Therefore, we cannot entirely exclude the possibility that a dimeric Alu element inserted prior to the divergence of prosimians and Anthropoidea, and that on the lineage leading to Anthropoidea, the dimeric element was replaced completely by a monomeric Alu element, or *vice versa*. Unlike in the other cited cases, however, replacement would have been complete. Furthermore, it would be necessary to postulate that the original or recombined Alu element was deleted in the lineage leading to tarsiers (see below). In conclusion, proposing two separate insertion events is more parsimonial, a rare but not unprecedented scenario.[70]

A second issue concerns the time of integration of the two Alu elements. Due to a small deletion around the integration site in tarsiers (Figure 2), we are unable to exclude the possibility that the FLAM-C-like BC200 predecessor integrated prior to the tarsier–anthropoid divergence and was later lost in tarsiers, after anthropoids had split off (Figure 8). This scenario would predict complete deletion of the FLAM-C-like element, including a few flanking nucleotides in tarsiers, a theoretical scenario we cannot completely rule out.

As the strepsirrhine G22 and L13 dimeric Alu elements appeared quite divergent, we wondered whether these two Alu families descend from a common ancestor or from independent progenitors. The integration of a dimeric Alu element in all strepsirrhine loci examined favours a single integration event prior to the strepsirrhine divergence. This is supported by the observation that the G22 and L13 Alu elements have at least six nucleotide substitutions in common, when compared to the AluJo and known lemuriform and lorisiform Alu consensus sequences (Figure 3). Thus, we suggest that the G22 and L13 Alu elements descend from a common ancestral AluJo-like element that must have integrated some time after strepsirrhines diverged from the common ancestor of primates, but before the Lemuriformes–Lorisiformes divergence. This is consistent with strepsirrhine monophyly documented recently by multi-locus and mono-locus markers.[71]

Although we realize that the tree reconstruction and bootstrap or puzzling support values are based on a relatively short sequence region (about 300 bp), its topology is highly congruent with the most accepted phylogenetic reconstructions in primates based on presence/absence data of SINEs.[72] For example, our phylogenetic reconstructions of the BC200/AluJo flanking regions (Figure 8) support the sister group relationship of the aye-aye and the rest of the Lemuriformes. In addition, numerous diagnostic positions in the Alu coding sequences also support monophyly of Lorisiformes and Lemuriformes, as well as that of the lorisiform families Galagidae and Lorisidae.[71]

Our sequence analyses further suggest that after the Lemuriformes and Lorisiformes divergence, the respective loci must have been subject to differential selective pressures. This is documented by the large discrepancy of conserved CpG doublets in the Alu coding regions (16–21 for Lorisiformes compared to only 9–11 for Lemuriformes), the presence of many indels and duplications in Lemuriformes, and the accumulation of mutations in the internal promoter elements in lemuriform Alu elements (Figure 3). An exception to this is the L13 Alu sequence of the aye-aye (*D. madagascariensis*) with 24 CpG dinucleotides and several lorisiform-like diagnostic AluJo mutations. This could be interpreted as a relatively close relationship of the aye-aye to Lorisiformes and/or earlier transcriptional activity that ceased in more recent evolutionary times (the aye-aye Alu element is not expressed in our transfection system). Another possibility is gene conversion (see above) with parts of a younger element, containing more CpG residues. However, only about ten CpG positions are shared with the expressed Alu elements of Lorisiformes; many of the additional CpG positions are at locations different from any of the known Alu consensus sequences, including those of prosimians (Figure 2). In any event, the example of the aye-aye Alu element illustrates that the mere count or density of CpG dinucleotides is not sufficient to predict transcriptional activity.

A novel dimeric Alu element transcribed in the *G. moholi* brain

We found that the *G. moholi* G22 Alu element is transcribed in the galago brain but not in three non-neural tissues examined, akin to BC200 RNA in anthropoids. In contrast, the *E. coronatus* L13 Alu element was transcriptionally silent (Figure 1). Further, all lorisiform G22 Alu elements derived from loci orthologous to the BC200 gene locus were effectively transcribed following transient transfection into HeLa cells, whereas the lemuriform L13 Alu elements were not (Figures 4 and 5). This raises the question of whether the lorisiform Alu elements became transcriptionally active after the strepsirrhine divergence, or whether the ancient dimeric Alu element was originally transcribed and became inactivated early in the lemuriform lineage, prior to the *Daubentonia* divergence. On the one hand, a newly integrated retronuon may at first be non-functional (termed a potonuon, as it has the potential to become active) and only later, after the acquisition of advantageous mutations, be recruited into a function (then termed a xaptonuon, as it was exapted into a function).[6] Recently, the high exaptative potential of previously silent, e.g. non-functional, Alu sequences as exons of protein-coding genes has been documented.[17,73,74] Thus, the ancient dimeric Alu element may have been silent for a long period of time before being transcribed. On the other hand, a smaller number of CpGs would then be conserved. Furthermore, it was postulated that a severe bottleneck occurred when the ancestors of Lemuriformes settled into

Madagascar from the African continent[75] about 47–80 million years ago.[71] Thus, it is equally conceivable that the ancient Alu element may have been inactivated in a common ancestor of all lemuriforms, while transcription was maintained in lorisiforms.

It is tempting to speculate that the G22 Alu RNA may have been co-opted into a function in the lorisiform central nervous system, analogous to anthropoid BC200 RNA and rodent BC1 RNA. BC200 and BC1 RNA have been proposed to play a role in regulation of dendritic protein synthesis.[76–78] Interestingly, targeted deletion of the BC1 RNA gene in mice results in behavioural changes best interpreted in terms of reduced exploration and increased anxiety.[79,80] Therefore, the recruitment of non-messenger RNAs into cellular functions may cause subtle behavioural changes, which in turn might be an evolutionary means to drive speciation events.[4,5] The differential transcriptional activity in the two Strepsirrhini branches is a serendipitous coincidence, as we can compare the relatively high level of sequence conservation in Lorisoidea with the rapid changes in Lemuroidea. Consequently, we can infer exaptation of the G22 Alu RNA in Lorisoidea into an unknown function, similar to BC200 RNA in Anthropoidea.

Upstream flanking sequences are required for transcription of the G22 Alu

We showed that 5′ flanking sequences are necessary for transcription of the *G. moholi* Alu element. Although highly conserved among Lorisiformes and Anthropoidea, deletion of the *G. moholi* DSE did not influence transcription of the Alu element when transiently transfected into HeLa cells. In contrast, the proximal flanking sequence including the TATA motif and the PSE were necessary for wild-type transcription levels of the G22 dimeric Alu element (Figure 6). Removal of the TATA-like sequence resulted in a complete loss of transcriptional activity. Therefore, the first 86 bp upstream sequence, and especially the TATA-like motif, was required for transcriptional activity of the *G. moholi* G22 Alu.

We could not observe transcription of the *E. coronatus*, *P. verreauxi* or *D. madagascariensis* Alu elements even when the respective Alu elements were under influence of the *G. moholi* 5′ flank (Figure 5, data not shown). This suggests that the lemuriform Alu elements are inactive due to alterations in internal promoter elements and/or non-functional terminators.[81,82] The requirement for appropriate 5′ flanking sequences is supported by the observation that the *P. verreauxi* 5′ flanking sequence was able to support transcription of the G22 Alu element. Interestingly, even the distantly related colugo locus could augment transcription of the G22 Alu, while the tarsier locus was only moderately active (Figure 7). The latter finding was somewhat surprising, as we expected the transcriptional activity of the G22 Alu including its TATA-like motif in the tarsier background to reach levels comparable to that of the *G. moholi* deletion mutant −22. Nonetheless, the data indicate that orthologous loci have maintained the potential to support transcription, even in species featuring empty target sites. Taken together, our transcriptional analyses suggest that both internal and external promoter elements are required for transcription of the G22 Alu element, once more supporting earlier findings with other retronuons,[35–37,39] and hopefully laying to rest the persisting misbelief that any Alu element, as long as it features intact internal promoter elements, can be transcribed *in vivo* under normal conditions.

In conclusion, the BC200 RNA gene locus served as an integration target for two distinct retronuons in its evolutionary history, a FLAM-C-like monomeric Alu element in anthropoids, resulting in the neural BC200 RNA transcript, and a dimeric AluJo-like element in strepsirrhines, giving rise to the lorisiform G22 and lemuriform L13 Alu elements. The *G. moholi* G22 Alu is transcribed in the galago brain, similar to its anthropoid analogue BC200 RNA. The significantly higher sequence conservation among the G22 Alu elements in comparison to the L13 orthologues indicates that the G22 Alu RNA has been exapted into a function, thus far unknown. This work shows that a locus can be targeted by retroposition at least twice in different lineages and corroborates the impact of local sequences on transcriptional activity of an inserted retronuon. Likewise, alterations internal to the retronuon determine its transcriptional fate in separate lineages.

Materials and Methods

Isolation of genomic DNA

Genomic DNA specimens and animal tissue samples from *Eulemur coronatus* (E.co), *Propithecus verreauxi* (P.ve), *Cheirogaleus medius* (C.me), *Lepilemur ruficaudatus* (L.ru), *Lepilemur dorsalis* (L.do), *Daubentonia madagascariensis* (D.ma), *Galago moholi* (G.mo), *Otolemur garnettii* (O.ga), *Nycticebus pygmaeus* (N.py), *Loris tardigradus* (L.ta), *Tarsius syrichta* (T.sy), *Tarsius bancanus* (T.ba) and *Cynocephalus variegatus* (C.va), were provided by the German Primate Center (Goettingen, Germany) and the American Museum of Natural History (New York, USA). Except for *D. madagascariensis*, all genomic DNA samples were obtained in lyophilized form. Genomic DNA of *D. madagascariensis* was isolated from liver using standard methods.[83]

PCR amplification

PCR primers flanking the BC200 RNA gene locus were designed on the basis of the previously published sequence alignment of *G. moholi* (AF314807), *E. coronatus* (AF314808), *T. syrichta* (AF314809) and the human loci (AF314806). PCR amplifications of lorisiform loci

orthologous to the anthropoid BC200 RNA gene locus were performed using the primers BCGALF3 (5′ AAGCA CAGGCTGGAGAAGCTAGATAN 3′) and BCGALR2 (5′ CATTCCTTTTAACAGGGAGCAN 3′). The corresponding lemuriform and tarsier loci were amplified with the primers BCGALF3 (see above) and BCGALR3 (5′ TGGTATAATTTTGCTCAAGGCCAN 3′), applying each primer at a concentration of 4 μM. The respective *Cynocephalus variegatus* locus was amplified using the oligonucleotides MS141 (5′ CCTGTTTAAATTCAAAGC CAAAAGACAC 3′) and MS142 (5′ AGATACTGGGTT GACCTGCCTTCAA 3′). All PCR reactions were performed in PCR buffer (50 mM Tris–HCl (pH 9.5), 20 mM $(NH_4)_2SO_4$, 1 mM DTT, 1.5 mM $MgCl_2$, 0.005% (w/v) NP-40, with or without 1 M betaine and 5% (v/v) DMSO), 200 μM each dNTP, 1 μM or 4 μM each primer, 200–500 ng of genomic DNA template and five units of Taq or Taq/Pfu polymerases mixture (five units of Taq, 1.25 units of Pfu) in a total volume of 50–100 μl. The following PCR cycles were performed: I, 96 °C for five minutes; II, 96 °C for 30 s; III, 50–60 °C for 30 s; IV, 67/72 °C for one minute/kb PCR product for 30–35 cycles followed by; V, 67/72 °C for five to ten minutes. PCR products were cloned into the pCR®II-Topo® vector using the TOPO TA Cloning kit (Invitrogen Ltd) and sequenced. Plasmid DNA was purified using the Plasmid Maxi Kit (QIAGEN).

Generation of recombinant plasmid constructs

G. moholi 5′ deletion constructs were generated by PCR (for oligonucleotides, see Supplementary Table 1). These constructs possessed gradually reduced 5′ flanking sequences, but were identical with the −270 construct with respect to the Alu element and the 3′ downstream sequence. Construct −242 lacks the forward primer sequence used to amplify the strepsirrhine loci and construct −207 was designed such that the octamer motif (GTTTAAAT) is absent. Further reduction of the 5′ flank generated construct −86, stopping immediately upstream of the PSE, construct −61, ceasing downstream of the PSE, and construct −41, in which only the TA-rich region was included in the sequence. Finally, construct −22 harboured the right part of the TA-rich motif only, while construct +1 contained vector-derived 5′ flanking sequence. All PCR products were cloned as described above, and plasmid inserts were verified by sequencing. Positive clones were designated as pGmo-x (x = 5′ flanking sequence).

Chimeric constructs between various lemuriform loci and the *G. moholi* locus were generated by 5′ overlapping PCR.[84] The constructs LF/GA (E.co 5′ flank fused to the G.mo Alu element), PF/GA (P.ve 5′ flank fused to the G.mo Alu element), GF/LA (G.mo 5′ flank fused to the E.co Alu element), GF/PA (G.mo Alu element fused to the P.ve Alu element), GF/DA (G.mo 5′ flank fused to the D.ma Alu element) were designed such that the 5′ flanking sequence of the corresponding locus was fused to the Alu coding sequence including the 3′ flanking sequence. In contrast, the construct CF/GA was generated by inserting the *G. moholi* Alu element including its own 3′ unique sequence ceasing downstream of the terminator sequence into the C.va locus, i.e. the 5′ and 3′ flanking sequence of the C.va locus was maintained. Construct TF/GA-22 was designed similarly, but with one minor difference. The *G. moholi* Alu element including the first 22 nt of 5′ flanking sequence was inserted into the T.ba locus. The oligonucleotides used to generate the chimera are listed in Supplementary Table 2.

DNA sequencing and sequence analyses

DNA sequencing was performed with the BigDye® Terminator Cycle Sequencing Ready Reaction Kit, version 2.0 (PE Applied Biosystems). Sequencing products were run on an *ABI Prism® 3700* capillary sequenator or an *ABI Prism® 377* DNA sequenator (Perkin Elmer). Sequence analyses and contig assembly were performed with the aid of the Lasergene software package (DNAStar Inc.). Sequence similarities among the loci obtained were then determined using the BLAST2 sequences program†,[85] followed by generation of multiple sequence alignments using Dialign2‡.[86] Subsequent editing and refinements of the alignments were performed manually with the sequence editor program Genedoc§.

Transient transfection of HeLa cells

HeLa cells were cultured in DMEM, 10% (v/v) fetal calf serum (GibcoBRL) at 37 °C in a 5% (v/v) CO_2 atmosphere. Transient DNA transfections of HeLa cells were conducted at 50–70% cell confluence with the lipotransfection reagent DAC-30™ (Eurogentec) according to the manufacturer's recommendations. All transfection experiments were performed in OptiMEM-1 (GibcoBRL) using 2 μg of plasmid constructs and 1 μg of control plasmid pCMV-MBII-52.[57] Total RNA was isolated 36–48 hours post-transfection.

Northern hybridization

Total RNA from HeLa cell cultures and animal tissues was isolated using the TRIzol reagent (Invitrogen Ltd). Human brain material was obtained from Dr Mehrain at the Brain Bank of the Institute for Neuropathology, Ludwig Maximilian University of Munich (Germany). *G. moholi*, *E. coronatus* and *T. syrichta* were obtained from the Duke University Primate Center, Durham, NC.

Total RNA was size-fractionated on a 7 M urea/8% (w/v) polyacrylamide gel and transferred to positively charged nylon membranes (BrightStar™-Plus, Ambion) using the semi-dry transfer procedure (Transblot SD, Bio-Rad). After UV-crosslinking, membranes were hybridized with ^{32}P-labelled oligonucleotides in 0.5 M sodium phosphate (pH 6.5 at 58 °C), 7% (w/v) SDS. Membranes were washed twice in 0.1 M sodium phosphate (pH 6.5 at 58 °C), 1% SDS, and then exposed to MS film (Kodak) at −80 °C using intensifying screens (Kodak).

For the detection of strepsirrhine Alu transcripts, several oligonucleotide probes were designed: Alu_5′ (5′ CCTCCCAGAGTGCT(AG)GGATTACAGGCGTGAGC CACC 3′; positions 44–8, *G. moholi* G22 sequence), G22_3′ (5′ TGTGCCCTGAGCAGAGGGCAGTGGCACCGGAG CCCAC 3′; positions 258–221, *G. moholi* G22 sequence), L13_3′ (5′ ACCGCCCTGGG-TAGGGTGCAGTGGCGC CATTGTCGC 3′; positions 289–253, *E. coronatus* L13 sequence) and L13_insert (5′ AGGTAGTCTCACT CTTGCTCAAGAGAGGGGGC 3′; positions 130–98,

† http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html.

‡ http://bioweb.pasteur.fr/seqanal/interfaces/dialign2.html.

§ http://www.psc.edu/biomed/genedoc.

E. coronatus L13 sequence). The MBII-52 snoRNA transcript was detected with the oligonucleotide MBII-52_P19 (5′ CCTCAGCGTAATCCTATTGAGCATGAA 3′). In addition, an oligonucleotide complementary to 7SL RNA was used (7SL_3′: 5′ AAGAGACGGGGTCTCGC TATG 3′).

Statistical analysis

For statistical analysis, Northern blots were densitized using the NIH Image software version 1.62. The intensities of the G22 Alu bands were normalized against those of the MBII-52 bands to correct for loading differences. The normalized intensities of the bands from the *G. moholi* −270 construct (Figure 6) or the *G. moholi* −86 construct (Figure 7) were set arbitrarily at 100% transcription. The relative transcription levels of all other constructs were determined as a percentage of that of the respective reference construct for each individual experiment performed. Student's *t*-test was performed using SigmaPlot 8.0.

Phylogenetic analyses

Phylogenetic reconstructions were performed for orthologous sequences of the BC200 locus in primates and *C. variegatus* as the primates' sister group.[55] The inserted FLAM in anthropoids and the AluJo in strepsirrhines were excluded from the analysis. The corresponding human sequence is 319 nt in length. Phylogenetic analyses were performed using maximum parsimony (MP; PAUP* 4.0b10;[87] branch-and-bound search), a distance based method implemented in PAUP* (NJ; HKY model of sequence evolution), maximum likelihood (ML; TREE-PUZZLE 5.0;[88] with the HKY model of sequence evolution)[89] approximating a gamma distribution of rates across sites by introducing four rate categories. All bootstrap and puzzling supports are based on 1000 replicates. The resulting tree topology and the range of bootstrap or puzzling supports are shown in Figure 8.

Acknowledgements

We thank Christian Roos for providing DNA samples and for comments on the manuscript. This work was supported by the German Human Genome Project through the BMBF (01KW9966) and by the Interdisciplinary Centre of Clinical Research (IZKF), Münster (Bro-3/054/04).

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.jmb.2005.03.058

References

1. Rogers, J. (1983). Retroposons defined. *Nature*, **301**, 460.
2. Rogers, J. H. (1985). The origin and evolution of retroposons. *Int. Rev. Cytol.* **93**, 187–279.
3. Weiner, A. M., Deininger, P. L. & Efstratiadis, A. (1986). Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* **55**, 631–661.
4. Brosius, J. (2005). Echoes from the past-are we still in an RNP world? *Cytogenet. Genome Res*. In the press.
5. Brosius, J. (1991). Retroposons-seeds of evolution. *Science*, **251**, 753.
6. Brosius, J. & Gould, S. J. (1992). On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". *Proc. Natl Acad. Sci. USA*, **89**, 10706–10710.
7. Britten, R. J. (1996). DNA sequence insertion and evolutionary variation in gene regulation. *Proc. Natl Acad. Sci. USA*, **93**, 9374–9377.
8. Brosius, J. (1999). Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica*, **107**, 209–238.
9. Brosius, J. (1999). RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene*, **238**, 115–134.
10. Brosius, J. & Tiedge, H. (1995). Reverse transcriptase: mediator of genomic plasticity. *Virus Genes*, **11**, 163–179.
11. Brosius, J. & Gould, S. J. (1993). Molecular constructivity. *Nature*, **365**, 102.
12. Britten, R. J. (1997). Mobile elements inserted in the distant past have taken on important functions. *Gene*, **205**, 177–182.
13. Kidwell, M. G. & Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. *Proc. Natl Acad. Sci. USA*, **94**, 7704–7711.
14. Makalowski, W. (2003). Genomics. Not junk after all. *Science*, **300**, 1246–1247.
15. Tomilin, N. V. (1999). Control of genes by mammalian retroposons. *Int. Rev. Cytol.* **186**, 1–48.
16. Nekrutenko, A. & Li, W. H. (2001). Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* **17**, 619–621.
17. Sorek, R., Ast, G. & Graur, D. (2002). Alu-containing exons are alternatively spliced. *Genome Res.* **12**, 1060–1067.
18. Courseaux, A. & Nahon, J. L. (2001). Birth of two chimeric genes in the Hominidae lineage. *Science*, **291**, 1293–1297.
19. Long, M., Betran, E., Thornton, K. & Wang, W. (2003). The origin of new genes: glimpses from the young and old. *Nature Rev. Genet.* **4**, 865–875.
20. DeChiara, T. M. & Brosius, J. (1987). Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. *Proc. Natl Acad. Sci. USA*, **84**, 2624–2628.
21. Martignetti, J. A. & Brosius, J. (1993). BC200 RNA: a neural RNA polymerase III product encoded by a monomeric Alu element. *Proc. Natl Acad. Sci. USA*, **90**, 11563–11567.
22. Okada, N. (1991). SINEs. *Curr. Opin. Genet. Dev.* **1**, 498–504.
23. Daniels, G. R. & Deininger, P. L. (1985). Repeat sequence families derived from mammalian tRNA genes. *Nature*, **317**, 819–822.
24. Ullu, E. & Tschudi, C. (1984). Alu sequences are processed 7SL RNA genes. *Nature*, **312**, 171–172.
25. Britten, R. J., Baron, W. F., Stout, D. B. & Davidson, E. H. (1988). Sources and evolution of human Alu repeated sequences. *Proc. Natl Acad. Sci. USA*, **85**, 4770–4774.

26. Jurka, J. & Smith, T. (1988). A fundamental division in the Alu family of repeated sequences. *Proc. Natl Acad. Sci. USA*, **85**, 4775–4778.
27. Quentin, Y. (1988). The Alu family developed through successive waves of fixation closely connected with primate lineage history. *J. Mol. Evol.* **27**, 194–202.
28. Maraia, R. J., Driscoll, C. T., Bilyeu, T., Hsu, K. & Darlington, G. J. (1993). Multiple dispersed loci produce small cytoplasmic Alu RNA. *Mol. Cell. Biol.* **13**, 4233–4241.
29. Kapitonov, V. & Jurka, J. (1996). The age of Alu subfamilies. *J. Mol. Evol.* **42**, 59–65.
30. Deininger, P. L., Batzer, M. A., Hutchison, C. A., III & Edgell, M. H. (1992). Master genes in mammalian repetitive DNA amplification. *Trends Genet.* **8**, 307–311.
31. Cordaux, R., Hedges, D. J. & Batzer, M. A. (2004). Retrotransposition of Alu elements: how many sources? *Trends Genet.* **20**, 464–467.
32. Matera, A. G., Hellmann, U. & Schmid, C. W. (1990). A transpositionally and transcriptionally competent Alu subfamily. *Mol. Cell. Biol.* **10**, 5424–5432.
33. Sinnett, D., Richer, C., Deragon, J. M. & Labuda, D. (1992). Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences. *J. Mol. Biol.* **226**, 689–706.
34. Shaikh, T. H., Roy, A. M., Kim, J., Batzer, M. A. & Deininger, P. L. (1997). cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *J. Mol. Biol.* **271**, 222–234.
35. Kramerov, D. A., Tillib, S. V., Shumyatsky, G. P. & Georgiev, G. P. (1990). The most abundant nascent poly(A)$^+$ RNAs are transcribed by RNA polymerase III in murine tumor cells. *Nucl. Acids Res.* **18**, 4499–4506.
36. Martignetti, J. A. & Brosius, J. (1995). BC1 RNA: transcriptional analysis of a neural cell-specific RNA polymerase III transcript. *Mol. Cell. Biol.* **15**, 1642–1650.
37. Chesnokov, I. & Schmid, C. W. (1996). Flanking sequences of an Alu source stimulate transcription *in vitro* by interacting with sequence-specific transcription factors. *J. Mol. Evol.* **42**, 30–36.
38. Roy, A. M., West, N. C., Rao, A., Adhikari, P., Aleman, C., Barnes, A. P. & Deininger, P. L. (2000). Upstream flanking sequences and transcription of SINEs. *J. Mol. Biol.* **302**, 17–25.
39. Arnaud, P., Yukawa, Y., Lavie, L., Pelissier, T., Sugiura, M. & Deragon, J. M. (2001). Analysis of the SINE S1 Pol III promoter from Brassica; impact of methylation and influence of external sequences. *Plant J.* **26**, 295–305.
40. Schmid, C. & Maraia, R. (1992). Transcriptional regulation and transpositional selection of active SINE sequences. *Curr. Opin. Genet. Dev.* **2**, 874–882.
41. Jurka, J. & Milosavljevic, A. (1991). Reconstruction and analysis of human Alu genes. *J. Mol. Evol.* **32**, 105–121.
42. Watson, J. B. & Sutcliffe, J. G. (1987). Primate brain-specific cytoplasmic transcript of the Alu repeat family. *Mol. Cell. Biol.* **7**, 3324–3327.
43. Skryabin, B. V., Kremerskothen, J., Vassilacopoulou, D., Disotell, T. R., Kapitonov, V. V., Jurka, J. & Brosius, J. (1998). The BC200 RNA gene and its neural expression are conserved in Anthropoidea (primates). *J. Mol. Evol.* **47**, 677685.
44. Kuryshev, V. Y., Skryabin, B. V., Kremerskothen, J., Jurka, J. & Brosius, J. (2001). Birth of a gene: locus of neuronal BC200 snmRNA in three prosimians and human BC200 pseudogenes as archives of change in the Anthropoidea lineage. *J. Mol. Biol.* **309**, 1049–1066.
45. Tiedge, H., Chen, W. & Brosius, J. (1993). Primary structure, neural-specific expression, and dendritic location of human BC200 RNA. *J. Neurosci.* **13**, 2382–2390.
46. Cheng, J. G., Tiedge, H. & Brosius, J. (1997). Expression of dendritic BC200 RNA, component of a 11.4S ribonucleoprotein particle, is conserved in humans and simians. *Neurosci. Letters*, **224**, 206–210.
47. Kremerskothen, J., Zopf, D., Walter, P., Cheng, J. G., Nettermann, M., Niewerth, U. *et al*. (1998). Heterodimer SRP9/14 is an integral part of the neural BC200 RNP in primate brain. *Neurosci. Letters*, **245**, 123–126.
48. Muddashetty, R., Khanam, T., Kondrashov, A., Bundman, M., Iacoangeli, A., Kremerskothen, J. *et al*. (2002). Poly(A)-binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles. *J. Mol. Biol.* **321**, 433–445.
49. West, N., Roy-Engel, A. M., Imataka, H., Sonenberg, N. & Deininger, P. (2002). Shared protein components of SINE RNPs. *J. Mol. Biol.* **321**, 423–432.
50. Basile, V., Vicente, A., Martignetti, J. A., Skryabin, B. V., Brosius, J. & Kennedy, J. L. (1998). Assignment of the human BC200 RNA gene (BCYRN1) to chromosome 2p16 by radiation hybrid mapping. *Cytogenet. Cell. Genet.* **82**, 271–272.
51. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J. *et al*. (2001). Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
52. Schmitz, J., Ohme, M. & Zischler, H. (2001). SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. *Genetics*, **157**, 777–784.
53. Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209–213.
54. Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucl. Acids Res.* **8**, 1499–1504.
55. Schmitz, J., Ohme, M., Suryobroto, B. & Zischler, H. (2002). The colugo (*Cynocephalus variegatus* Dermoptera): the primates' gliding sister? *Mol. Biol. Evol.* **19**, 2308–2312.
56. Zuker, M. (2003). MFOLD web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* **31**, 3406–3415.
57. Cavaillé, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C. I., Horsthemke, B. *et al*. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci. USA*, **97**, 14311–14316.
58. Chen, Y., Sinha, K., Perumal, K., Gu, J. & Reddy, R. (1998). Accurate 3′ end processing and adenylation of human signal recognition particle RNA and alu RNA *in vitro*. *J. Biol. Chem.* **273**, 35023–35031.
59. Pelissier, T., Bousquet-Antonelli, C., Lavie, L. & Deragon, J. M. (2004). Synthesis and processing of tRNA-related SINE transcripts in *Arabidopsis thaliana*. *Nucl. Acids Res.* **32**, 3957–3966.
60. Ullu, E. & Weiner, A. M. (1985). Upstream sequences modulate the internal promoter of the human 7SL RNA gene. *Nature*, **318**, 371–374.
61. Kunkel, G. R. (1991). RNA polymerase III transcription of genes that lack internal control regions. *Biochim. Biophys. Acta*, **1088**, 1–9.
62. Myslinski, E., Ame, J. C., Krol, A. & Carbon, P. (2001).

An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene. *Nucl. Acids Res.* **29**, 2502–2509.
63. Schaub, M., Myslinski, E., Schuster, C., Krol, A. & Carbon, P. (1997). Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III. *EMBO J.* **16**, 173–181.
64. Jurka, J. & Klonowski, P. (1996). Integration of retroposable elements in mammals: selection of target sites. *J. Mol. Evol.* **43**, 685–689.
65. Dewannieux, M., Esnault, C. & Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nature Genet.* **35**, 41–48.
66. Schmitz, J., Churakov, G., Zischler, H. & Brosius, J. (2004). A novel class of mammalian-specific tailless retropseudogenes. *Genome Res.* **14**, 1911–1915.
67. Feng, Q., Moran, J. V., Kazazian, H. H., Jr & Boeke, J. D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905–916.
68. Tatout, C., Lavie, L. & Deragon, J. M. (1998). Similar target site selection occurs in integration of plant and mammalian retroposons. *J. Mol. Evol.* **47**, 463–470.
69. Kass, D. H., Batzer, M. A. & Deininger, P. L. (1995). Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. *Mol. Cell. Biol.* **15**, 19–25.
70. Kass, D. H., Raynor, M. E. & Williams, T. M. (2000). Evolutionary history of B1 retroposons in the genus Mus. *J. Mol. Evol.* **51**, 256–264.
71. Roos, C., Schmitz, J. & Zischler, H. (2004). Primate jumping genes elucidate strepsirrhine phylogeny. *Proc. Natl Acad. Sci. USA*, **101**, 10650–10654.
72. Schmitz, J., Roos, C. & Zischler, H. (2005). Primate phylogeny: molecular evidence from retroposons. *Cytogenet. Genome Res.* **108**, 26–37.
73. Lev-Maor, G., Sorek, R., Shomron, N. & Ast, G. (2003). The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons. *Science*, **300**, 1288–1291.
74. Singer, S. S., Männel, D. N., Hehlgans, T., Brosius, J. & Schmitz, J. (2004). From "junk" to gene: curriculum vitae of a primate receptor isoform gene. *J. Mol. Biol.* **341**, 883–886.
75. Go, Y., Satta, Y., Kawamoto, Y., Rakotoarisoa, G., Randrianjafy, A., Koyama, N. & Hirai, H. (2002). Mhc-DRB genes evolution in lemurs. *Immunogenetics*, **54**, 403–417.
76. Wang, H., Iacoangeli, A., Popp, S., Muslimov, I. A., Imataka, H., Sonenberg, N. *et al.* (2002). Dendritic BC1 RNA: functional role in regulation of translation initiation. *J. Neurosci.* **22**, 10232–10241.
77. Brosius, J. & Tiedge, H. (1995). Neural BC1 RNA: dendritic localization and transport. In *Localized RNAs* (Lipshitz, H. D., ed.), pp. 289–330, R.G. Landes, Austin, TX.
78. Brosius, J. & Tiedge, H. (2001). Neuronal BC1 RNA: intracellular transport and activity-dependent modulation. *Results Probl. Cell. Differ.* **34**, 129–138.
79. Skryabin, B. V., Sukonina, V., Jordan, U., Lewejohann, L., Sachser, N., Muslimov, I. *et al.* (2003). Neuronal untranslated BC1 RNA: targeted gene elimination in mice. *Mol. Cell. Biol.* **23**, 6435–6441.
80. Lewejohann, L., Skryabin, B. V., Sachser, N., Prehn, C., Heiduschka, P., Thanos, S. *et al.* (2004). Role of a neuronal small non-messenger RNA: behavioural alterations in BC1 RNA-deleted mice. *Behav. Brain Res.* **154**, 273–289.
81. Maraia, R. J., Chang, D. Y., Wolffe, A. P., Vorce, R. L. & Hsu, K. (1992). The RNA polymerase III terminator used by a B1-Alu element can modulate 3′ processing of the intermediate RNA product. *Mol. Cell. Biol.* **12**, 1500–1506.
82. Chu, W. M., Liu, W. M. & Schmid, C. W. (1995). RNA polymerase III promoter and terminator elements affect Alu RNA expression. *Nucl. Acids Res.* **23**, 1750–1757.
83. Sambrook, J., Fritsch, E. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual* (2nd edit.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
84. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, **77**, 61–68.
85. Tatusova, T. A. & Madden, T. L. (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Letters*, **174**, 247–250.
86. Morgenstern, B. (1999). DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics*, **15**, 211–218.
87. Swofford, D.L. (2000). PAUP*: Phylogenetic Analysis using Parsimony (*and other Methods) Version 4, Sinauer Associates, Sunderland, MA.
88. Strimmer, K. & von Haeseler, A. (1996). Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**, 964–969.
89. Hasegawa, M., Kishino, H. & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174.

Edited by J. Karn

(Received 30 January 2005; received in revised form 18 March 2005; accepted 21 March 2005)