Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 349 (2005) 197–205
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

PCR-based approach to SINE isolation: Simple and complex SINEs

Olga R. Borodulina, Dmitri A. Kramerov*

Laboratory of Eukaryotic Genome Evolution, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow 119991, Russia

Received 1 October 2004; received in revised form 19 November 2004; accepted 22 December 2004
Available online 14 March 2005
Received by S.M. Mirkin

Abstract

Highly repeated copies of short interspersed elements (SINEs) occur in eukaryotic genomes. The distribution of each SINE family is usually restricted to some genera, families, or orders. SINEs have an RNA polymerase III internal promoter, which is composed of boxes A and B. Here we propose a method for isolation of novel SINE families based on genomic DNA PCR with oligonucleotide identical to box A as a primer. Cloning of the size-heterogeneous PCR-products and sequencing of their terminal regions allow determination of SINE structure. Using this approach, two novel SINE families, Rhin-1 and Das-1, from the genomes of great horseshoe bat (*Rhinolophus ferrumequinum*) and nine-banded armadillo (*Dasypus novemcinctus*), respectively, were isolated and studied. The distribution of Rhin-1 is restricted to two of six bat families tested. Copies of this SINE are characterized by frequent internal insertions and significant length (200–270 bp). Das-1 being only 90 bp in length is one of the shortest SINEs known. Most of Das-1 nucleotide sequences demonstrate significant similarity to alanine tRNA which appears to be an evolutionary progenitor of this SINE. Together with three other known SINEs (ID, Vic-1, and CYN), Das-1 constitutes a group of simple SINEs. Interestingly, three SINE families of this group are alanine tRNA-derived. Most probably, this tRNA gave rise to short and simple but successful SINEs several times during mammalian evolution.

Keywords: Short retroposons; Alanine tRNA; Method; Rhinolophidae; Chiroptera; Dasypodidae; Edentata

1. Introduction

Short interspersed elements (SINEs) or short retroposons are 80–400 bp repetitive DNA sequences that proliferate in eukaryotic genomes via transcription followed by reverse transcription (Okada, 1991; Deininger and Batzer, 2002). There are usually one, two, or several SINE families in a genome. Each family consists of 10^3–10^5 SINE copies and their sequences being significantly different from each other (5–35%) depending on the time when these copies emerged. Over 30 SINE families have been isolated from genomes of mammals, reptiles, fishes, molluscs, ascidia, insects, and flowering plants. SINEs are transcribed by RNA polymerase III due to the presence of the internal promoter for RNA polymerase III in their 5′ region. This promoter is composed of two boxes (A and B) spaced by a 30–40 bp sequence. Most of SINE families are believed to have been derived from tRNA molecules because 5′ region of these genetic elements demonstrates reasonable sequence similarity to certain tRNAs. This region is usually followed by a unique sequence characterizing of each SINE family. A 3′end part of SINEs is terminated either by an A-rich tail (in mammals) or by short tandem repeats (in fish, reptiles, and invertebrates).

Isolation and studying of novel SINE families are important in several aspects. (1) Although it is known that SINEs are widely spread among eukaryotes (Okada and Ohshima, 1995; Borodulina and Kramerov, 1999), it is not clear how ubiquitous they are. There are eukaryotic organisms (yeast, *Drosophila*, or the nematode *C. elegans*) that have no SINEs in their genomes (Eickbush and Furano, 2002). Therefore, a question arises—what are the determi-

Abbreviations: SINE; Short interspersed element; LINE; Long interspersed element; LTR; Long terminal repeats.

* Corresponding author. Tel.: +7 95 135 9997; fax: +7 95 135 1405.
E-mail address: kramerov@eimb.ru (D.A. Kramerov).

0378-1119/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.gene.2004.12.035

nants of SINEs distribution among eukaryotes? Particularly, we are interested to know whether all mammals, and others vertebrates, have SINEs in their genome. Answers to these questions may shed light on principles of SINE origin, and allow us to understand SINE's impact on host-organisms. (2) The study of novel SINE families may contribute to the clarification of principles of structural organization, functioning, and evolution of these genetic elements. (3) Individual copies and entire SINE families proved to be excellent phylogenetic markers (Shimamura et al., 1997; Stoneking et al., 1991; Serdobova and Kramerov, 1998; Kramerov et al., 1999; Shedlock and Okada, 2000; Schmitz et al., 2002; Nikaido et al., 2003; Roos et al., 2004), therefore isolation of SINEs from species that have not yet been studied may be very useful for the investigation of the phylogenetic relationship between organisms.

To search and isolate novel SINE families, we previously developed a simple method of SINE detection and cloning that we designated (AB)-PCR (Borodulina and Kramerov, 1999, 2001). In this approach, a polymerase chain reaction (PCR) was carried out with very small amount of genomic DNA as a template and two oligonucleotides specific to boxes A and B of the promoter of RNA polymerase III as primers. The resulting PCR-product containing a 30–40 bp sequence located between boxes A and B of SINEs was used as a probe for screening a genomic library. Although this approach allows easy finding of the clones containing SINEs, determination of SINE nucleotide sequences requires a labor-consuming subcloning of DNA insertions and extensive sequencing experiments.

In this work, in order to simplify cloning and sequencing of SINEs, we developed another PCR-based method for isolation of novel SINE families and determination of their structure. This approach designated as (A)-PCR includes PCR of genomic DNA with only one primer, which is identical to the A box of RNA polymerase III promoter. In such a reaction, the genomic regions containing SINE copies that are "tail-to-tail" oriented and located rather close to each other can be amplified (Fig. 1, Stage I). The heterogeneous PCR-products are cloned and the terminal regions of DNA insertions are sequenced. This allows determination of the SINE structure, except for its 5′region. Then PCR of genomic DNA with a primer complementary to the consensus of 3′region of the SINE family is carried out (Fig. 1, Stage II). Subsequent cloning and sequencing of the ends of amplified DNA fragments enable determination of SINE nucleotide sequence including the 5′region. Using this approach, we isolated and described two novel SINEs families: Das-1 from the genome of nine-banded armadillo and Rhin-1 from the genome of great horseshoe bat. Das-1 seems to be one of the shortest known SINE. Together with several other SINE families, Das-1 forms a group of simple SINEs.

2. Materials and methods

2.1. Tissue and DNA samples

Bat tissues were provided by Drs. V. Matveev and A. Borisenko (Moscow State University, Russia). Liver sample of tree kangaroo was obtained from Moscow Zoo. Nine-banded armadillo and rock lizard DNAs were donated by Drs. R. DeBry (University of Cincinnati, USA) and V. Grechko (Institute of Molecular Biology, Moscow, Russia), respectively.

DNA was isolated from fresh or ethanol-fixed tissues by incubation with proteinase K followed by phenol/chloroform extraction.

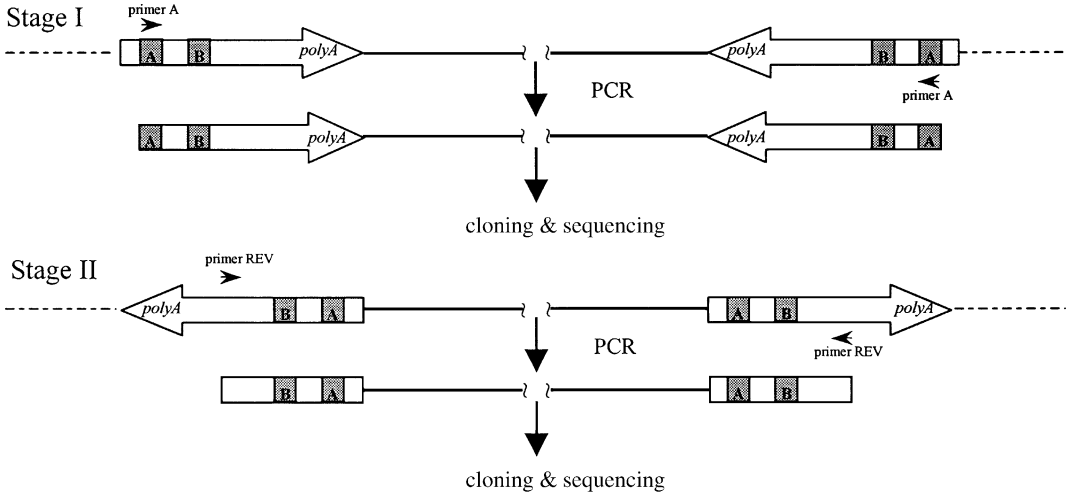

Fig. 1. Scheme of PCR-based isolation of SINEs. Stage I: (A)-PCR. A hypothetical genomic region with two "tail-to-tail" oriented SINE copies (boxed) is shown. A and B boxes of the RNA polymerase III promoter are shaded. PCR with a primer identical to box A (primer A) results in amplification of the region. Cloning of the PCR-product and sequencing of its terminal regions allow determination of the nucleotide sequence of a large part of the corresponding SINE. Stage II: "Reverse" PCR. A hypothetical genomic region with two "head-to-head" oriented SINE copies is shown. PCR with a primer specific to 3′region of the SINE amplifies the given region. Cloning and sequencing of the PCR product reveal the structure of the 5′region of the SINE.

2.2. PCR methods

(AB)-PCR was carried out as described previously (Borodulina and Kramerov, 1999) with modifications. In brief, the reaction mixture (100 μl) contained 0.1 ng of genomic DNA and two 11-nucleotide primers ("A": 5′-TRGCTCAGTGG-3′ and "B": 5′-GGRATYGAACY-3′) specific for A and B boxes of RNA pol III promoter consensus, respectively. After 27 PCR cycles (95 °C, 1 min; 30 °C, 1 min; 72 °C, 1 min), the amplified ~55 bp DNA fragments were isolated by electrophoresis in 5% NuSieve (FMC) agarose gel followed by electroelution on DEAE membrane (NA 45, Schleicher and Schuell).

(A)-PCR was performed according to the same protocol but 10 ng genomic DNA and the only primer (A) were used. After separation in agarose gel, PCR products longer than 200 bp were collected by reverse electrophoresis on DEAE membrane placed into the loading wells.

"Reverse" PCR was similar to (A)-PCR except for using one of the following primers: DnoRev1 (annealing 38 °C; see Fig. 3), RfeRev1, or RfeRev2 (annealing 50 °C; see Fig. 4).

Rhin-1 SINE was detected in the genomes of various bats using 25 PCR cycles (95 °C, 1 min; 50 °C, 1 min; 72 °C, 1 min) with 0.1 or 10 ng genomic DNAs and primers RfeForw1 and RfeRev1 (see Fig. 4).

For radioactive labeling, 0.5% of isolated PCR product was added to a PCR reaction mix (25 μl) containing 30 μCi [α-^{32}P]dATP and 20 PCR cycles were carried out.

2.3. Library construction and screening

Half of the isolated (A)-PCR product was cloned into pGEM-T Vector (Promega) according to the manufacturer's instructions. Colony hybridization was carried out at 60 °C in 4×SSC, 0.5% SDS, 5×Denhardt's solution, 0.1 mg/ml boiled herring sperm DNA, and ^{32}P-labeled (AB)-PCR product. Filters were washed in 0.1% SSC, 0.1% SDS at 42 °C, and positive colonies were detected by autoradiography.

For isolation of full-size Rhin-1 SINE copies, *Rhinolophus ferrumequinum* genomic library was prepared by ligation of genomic DNA digested with *Eco*RI and *Hin*dIII into pBluescript SK+(Stratagene). Screening of the library was performed using labeled PCR product obtained with Rhin-1 specific primers RfeForw1 and RfeRev1.

2.4. DNA sequencing

SINE copies were sequenced using dideoxynucleotide method with Sequenase 2.0 (Amersham Bioscience) and standard (forward and reverse) M13 primers according to the manufacturer's instructions with modification (Redston and Kern, 1994).

2.5. Dot-blot hybridization

Bat genomic DNAs (1 μg) were incubated in 10 μl of 0.5 M NaOH for 1 h at 37 °C and 20 volumes of 6×SSC, 6% formaldehyde, and 0.025 M NaH_2PO_4 were added. DNA was applied to a Hybond N membrane using a dot-blot apparatus. Hybridization and washing conditions as well as the probe were the same as in colony hybridization screening of the horseshoe bat genome library.

3. Results

3.1. (A)-PCR

The results of (A)-PCR and (AB)-PCR of genomic DNA from four mammalian species and a lizard are shown in Fig. 2. The expected ~55 bp DNA fragments were amplified in the (AB)-PCR of mouse (*Mus musculus*), nine-banded armadillo (*Dasypus novemcinctus*), great horseshoe bat (*R. ferrumequinum*), and rock lizard (*Darevskia* sp.) DNA. In the case of tree kangaroo (*Dendrolagus bennettianus*), the (AB)-PCR product was not observed that may be due to a deviation from canonical sequences in A and/or B boxes of presumptive kangaroo SINEs. When only a primer corresponding to the box A of RNA polymerase III promoter was used ((A)-PCR), the reaction products appeared as a smear, for all the species (Fig. 2). One can assume that the size-heterogeneous DNA fragments resulted from amplification of those genomic regions where two "tail-to-tail" oriented SINE copies were located not far from each other (up to ~3000 bp).

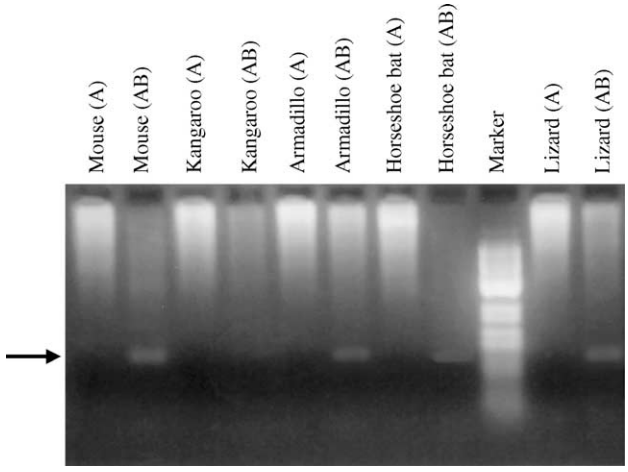

Fig. 2. Results of (A)- and (AB)-PCR of genomic DNAs from house mouse (*Mus musculus*), tree kangaroo (*Dendrolagus bennettianus*), nine-banded armadillo (*Dasypus novemcinctus*), great horseshoe bat (*Rhinolophus ferrumequinum*), and rock lizard (*Darevskia* sp.). Amplified DNA fragments were separated by electrophoresis in 5% agarose gel and stained with ethidium bromide. 55 bp (AB)-PCR-products are marked by an arrow.

The *D. novemcinctus* and *R. ferrumequinum* (A)-PCR products were cloned in a plasmid vector, and the libraries were screened by hybridization with labeled (AB)-PCR products obtained from corresponding species. The end regions of DNA insertions from positive clones were sequenced that allowed the identification of the two novel SINE families. This approach also allowed us to isolate a new SINE family from the rock lizard genome whose structure will be reported elsewhere.

3.2. Das-1, a SINE family from the armadillo genome

Fig. 3 shows the aligned nucleotide sequences of the terminal parts of the cloned *D. novemcinctus* (A)-PCR products (11 upper sequences). These sequences are about 80 bp long and contain box B, an A-rich tail, and numerous single-nucleotide differences, which is typical for SINE copies. These features suggest that the similar sequences are SINE copies. In order to establish the full structure of the SINE, PCR with *D. novemcinctus* genomic DNA and a primer complementary to a region in the 3' part of the presumptive SINE (primer DnoRev1) was carried out. This reaction resulted in products heterogeneous in length that looked like a smear after electrophoresis in agarose gel (data not shown). These products seem to correspond to DNA fragments flanked with SINE copies that are "head-to-head" oriented. After cloning of the PCR-products and sequencing of their end regions, we aligned these sequences *inter se* and with (A)-PCR products (Fig. 3). It allowed us to deduce the full structure of the SINE named Das-1. Search in nucleotide databases revealed two *D. novemcinctus* genomic sequences (AC145506 and AC149030) containing six non-annotated copies of Das-1, five of which are shown in Fig. 3. The presence of direct repeats flanking each of the copies (underlined) and their localization indicate that the deduced consensus corresponds to the full-size SINE. The length of Das-1 including an A-rich tail is only 90–100 bp that allows assessing it as one of the shortest SINEs known. The first 76 bp of Das-1 demonstrates a very significant similarity (82%) to alanine tRNA indicating the Das-1 origin from this tRNA. Two SINE families similar to Das-1 are known: rodent ID (Kim et al., 1994) and Vic-1 from camelids (Lin et al., 2001)(Fig. 3B). Especially high similarity (about 82% for their tRNA-related parts) was found between Das-1 and Vic-1. Beside the similarity to alanine tRNA, all these three SINE families share the small length and the fact that their tRNA-related region is directly followed by an A-rich tail (notice that short pyrimidine-containing sequences are present in the A-rich tail of ID and Vic-1, but not Das-1). Judging from the distribution among mammals (see Discussion), one can assume these three SINE families arose independently from each other. Alanine tRNA may be prone to production of SINEs characterized by the very short length, the lack of a tRNA-unrelated part, and the rather high level of nucleotide sequence similarity to tRNA.

3.3. Rhin-1, a SINE family from the horseshoe bat genome

The alignment of end parts of nucleotide sequences of cloned *R. ferrumequinum* (A)-PCR products (9 upper sequences) is shown in Fig. 4. In this case, the aligned sequences were significantly longer (about 180–255 bp) than they were in *D. novemcinctus*. There were very similar insertions of 72 to 78 bp long located upstream from the tail of the presumptive SINE in three clones (Rfe5rev, Rfe2Adir, and Rfe1Arev). "Reverse" PCR was carried out in two variants: with either RfeRev1 or RfeRev2 primers that were complementary to regions located before and after the 72–78 bp insertion, respectively. The end parts of the five cloned PCR-products obtained in this experiment are shown in Fig. 4. Their alignment allowed us to determine the 5′ end structure of the novel SINE which was named Rhin-1 (Fig. 4).

Considering the complex structure of Rhin-1, we verified the data obtained from PCR by sequencing of several full genomic copies cloned by the traditional method (see Materials and methods). In clones Rfe10f and Rfe13f, SINE copies well correspond to a Rhin-1 consensus, and the position of the direct flanking repeats (underlined) confirmed the correct identification of Rhin-1 ends (Fig. 4). A SINE copy in clone Rfe5dir had approximately the same level of similarity to a Rhin-1 consensus but it lacked a 50 bp sequence in its 3′ end part. The presence of an A-rich tail and flanking repeats (underlined) in this copy suggests that it is not truncated and probably may be a member of a specific Rhin-1 subfamily.

It should be noted that Rhin-1 apparently tends to the formation of additional internal nucleotide sequences. Besides the above-mentioned 72–78 bp insertion, two other copies of Rhin-1 (clones Rfe33dir and Rfe10f) contained long insertions in similar sites of the SINE. Both insertions were eliminated from the aligned copies but their location is marked by arrows in Fig. 4. We attempted to align these internal insertions (two bottom sequences) with nucleotide sequences of Rhin-1 (Fig. 4). A 47 bp insertion in Rfe33dir copy proved to be similar to the adjacent region, i.e. it apparently resulted from a tandem duplication of this region of Rhin-1. On the other hand, 102 bp insertion in copy Rfe10f showed a very limited similarity to a Rhin-1 sequence and its origin is not clear. Internal duplications and insertions often contribute to generation of new SINE subfamilies (Borodulina and Kramerov, 2001; Vassetzky et al., 2003). Likewise, Rhin-1 copies (Rfe5rev, Rfe2Adir, Rfe1Arev, Rfe33, and Rfe10) containing duplications and insertions can represent particular SINE subfamilies.

The analysis of Rhin-1 consensus revealed 70% sequence similarity between the 5′end region of this SINE and isoleucine tRNA suggesting the Rhin-1 origin from this tRNA (Fig. 4). The tRNA-unrelated region in Rhin-1 is rather long—at least 90 bp. Its end adjacent to an A-rich tail contains short alternate C and T blocks. Some features of the tRNA-unrelated region suggest that Rhin-1 probably con-

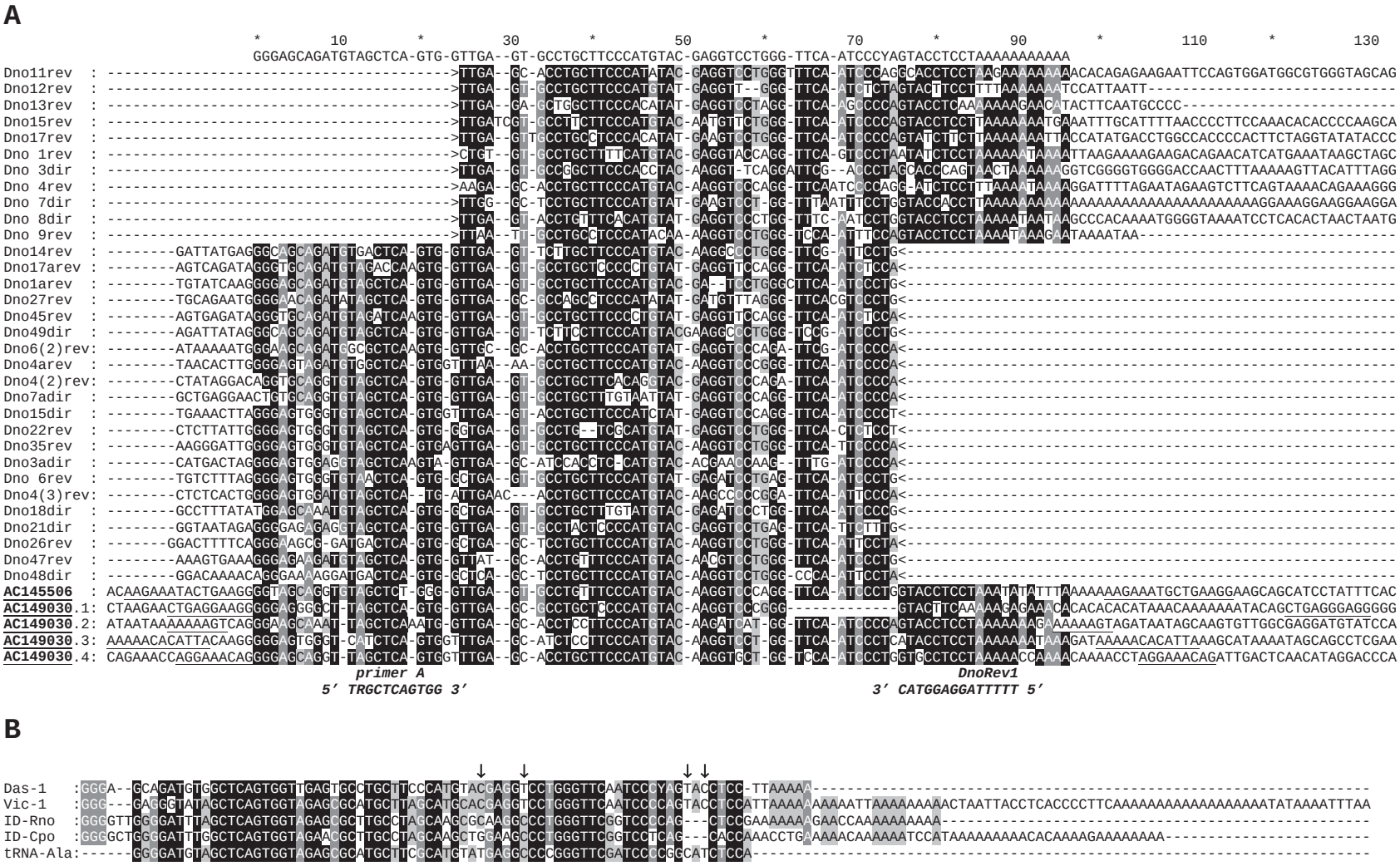

Fig. 3. Das-1, an armadillo SINE family. (A) Alignment of armadillo PCR-products that were obtained using primer A (11 upper sequences) or primer DnoRev1 (21 lower sequences). The primer sequences are replaced by the following marks:>, primer A and <, primer DnoRev1. GenBank accession numbers AY569616–AY569647. The bottom five sequences are SINE copies found in two *Dasypus novemcinctus* genomic fragments (AC145506 and AC149030) in nucleotide database. Direct repeats flanking each of this copy of Das-1 are underlined. (B) Alignment of consensus sequences of Das-1, Vic-1 (Lin et al., 2001), rat ID (subtype 2), guinea pig ID (Kim et al., 1994), and mouse $tRNA^{Ala(CGC)}$ (Russo et al., 1986). Arrows indicate parallel changes in Das-1 and Vic-1 relative to $tRNA^{Ala}$ unrelated to hypervariable CpG sites in tRNA.

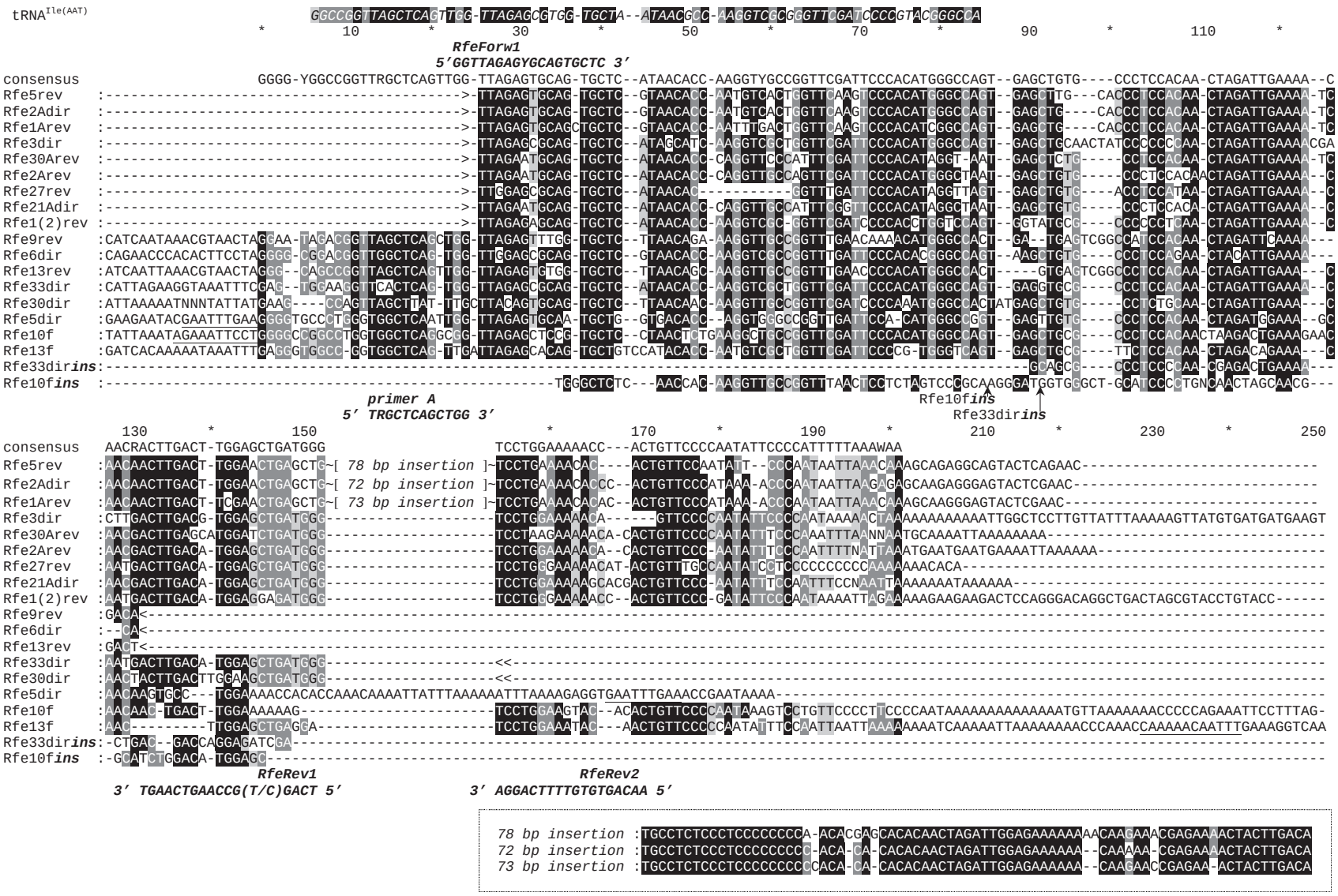

Fig. 4. Rhin-1, a horseshoe bat SINE family. Alignment of horseshoe bat PCR-products that were obtained using primers A (9 upper sequences), RfeRev1, or RfeRev2 (3 and 2 lower sequences, respectively) is shown. Sequences of the primers are replaced by the following marks: >(primer A), <(primer RfeRev1), and <<(primer RfeRev2). Three Rhin-1 copies (Rfe5dir, Rfe10f, and Rfe13f) were cloned from the *Rhinolophus ferrumequinum* genome by a traditional method (see Materials and methods). Direct repeats flanking these copies are underlined. The bottom two lines: additional sequences (insertions) in Rfe33dir and Rfe10f copies. These two insertions were removed from the Rhin-1 copies to maintain alignment. Location of the insertions is shown by arrows. Three additional sequences (78, 72, and 73 bp insertions) were removed from Rfe5rev, Rfe2Adir, and Rfe1Arev copies and their alignment is shown in a frame. GenBank accession numbers AY568167–AY568183. Mouse isoleucine tRNA$^{Ile(AAT)}$ (Russo et al., 1987) is shown at the top of the figure.

sists of two parts (monomers). There is an A-rich sequence (AAAACAA) in the tRNA-unrelated region. This sequence resembles an oligo(A) linker between monomers in such dimeric SINEs as Alu, MEN, or B1-dID (Deininger et al., 1981; Serdobova and Kramerov, 1998; Kramerov and Vassetzky, 2001). It should be noticed that an A-rich tail in the Rfe5dir copy coincides with the "linker", i.e. this copy consists of only the first of the two presumptive monomers. Perhaps Rfe5dir is a member of a monomeric SINE subfamily that can be an evolutionary precursor of the full-size Rhin-1.

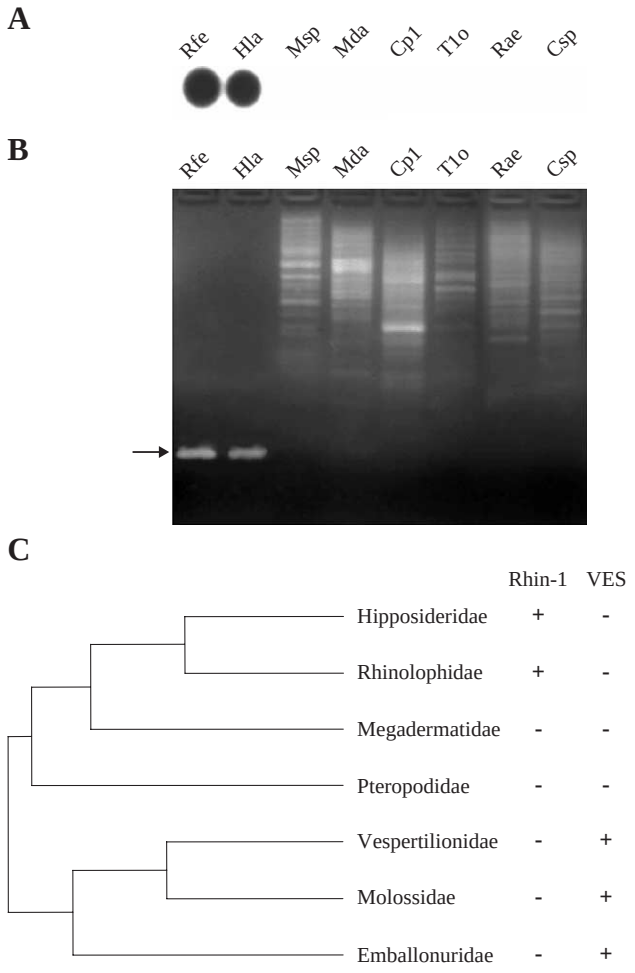

Fig. 5. Taxonomic distribution of the Rhin-1 SINE family. (A) Dot-hybridization of the labeled Rhin-1-specific probe to the genomic DNA from bats. Each dot contains 1 μg of genomic DNA. (B) PCR-based detection of the Rhin-1 in genomic bat DNAs. PCR was carried out with 0.1 ng of *Rhinolophus ferrumequinum* and *Hipposideros larvatus* DNAs or 10 ng of DNAs from other species. The expected DNA fragments (140 bp) are indicated by arrow. Rfe—*Rhinolophus ferrumequinum* (Rhinolophidae); Hla—*Hipposideros larvatus* (Hipposideridae); Msp—*Megaderma spasma* (Megadermatidae); Mda—*Myotis daubentoni* (Vespeptilionidae); Cpl—*Chaerephon plicata* (Mollosidae); Tlo—*Taphozous longimanus* (Emballonuridae); Rae—*Rousettus aegyptiacus* (Pteropodidae); and Csp—*Cynopterus sphinx* (Pteropodidae). (C) Simplified evolutionary tree of Chiroptera and distribution of SINE families. The tree is based on the data of Teeling et al. (2002) but includes only the chiropteran families where the presence of Rhin-1 and VES SINEs was tested. VES distribution is presented according to Borodulina and Kramerov (1999) and Kawai et al. (2002).

In order to explore which bats have Rhin-1 SINE in their genomes, dot-hybridization was performed. The DNA fragment amplified by PCR with primers RfeDir1 and RfeRev1 (see Fig. 4) was used as a probe. Hybridization signals were observed for representatives of the Rhinolophidae and Hipposideridea families but not for five other Chiroptera families tested (Fig. 5A). These data were supported by PCR of the genomic DNAs with primers RfeDir1 and RfeRev1. The PCR-product of the expected length was detected for rhinolophids and hipposiderids, whereas for bats of other families this PCR-product was not observed even when 100-fold excess (10 ng) of the genome DNA was used (Fig. 5B). The results show the presence of Rhin-1 SINE only in the rhinolophid and hipposiderid genomes, and that the emergence of this SINE was followed by the divergence these two bat families.

4. Discussion

4.1. PCR-based approach to the SINEs isolation

In this work we developed a method for isolation of SINEs based on PCR of genomic DNA with a primer which is identical to the box A of the RNA polymerase III promoter ((A)-PCR). This reaction is similar to inter-SINE-PCR, i.e. PCR with a primer (or primers) specific to a particular SINE family that results in amplification of genomic regions flanked by SINE copies. Previously, inter-SINE-PCR was used to detect known SINEs in various organisms (Jurka et al., 1995), to study phylogenetic relationships between species (Bannikova et al., 2002), and to map the genome loci (Kass and Batzer, 1995). Unlike inter-SINE-PCR, (A)-PCR was developed for isolation of novel SINE families. The ability of this reaction to amplify the genome regions containing SINEs with unknown nucleotide sequence is based on rather high conservation of the box A. However, in contrast to what one might expect (Fig. 1), in almost all clones, the nucleotide sequences of SINEs were detected at only one end of the (A)-PCR product. This may be due to the short length (11 bp) of the primer A and its degenerate nucleotide sequence which may result in occasionally binding of the primer A to non-SINE genomic sequences during PCR. Such non-SINE sequences can include either random A box-like sites and numerous tRNA tailless retropseudogenes (Schmitz et al., 2004). Although this feature of the (A)-PCR reduces the yield of the method, instead it allows amplification of detached SINE copies and, hence, should work with moderately repeated SINEs.

The method developed in this work has the following advantages. First, it enables a purposeful search for SINEs in the genomes of not yet studied species. In other words, it helps to avoid cloning and sequencing of other repetitive

sequences: satellite and minisatellite DNAs, LINEs, LTR-containing retrotransposons, and DNA transposons, which is typical for the widely used method based on the hybridization screening of a genomic library by the labeled total genomic DNA. Second, this method does not require the hybridization search for a SINE within a cloned insertion, as well as its subcloning. Third, the volume of sequencing decreases dramatically and "long reading" becomes unnecessary because the sequences of interest are located at the ends of the cloned (A)-PCR products. Fourth, this method allows one to use a very small amount of genomic DNA, which is sometimes critical (for example, only 200 ng DNA of armadillo was available to us).

Some more advantages of the proposed method can be noted. Our previously published method (AB)-PCR (Borodulina and Kramerov, 1999) requires good matching of two primers (A and B) to A and B boxes in SINEs. However, there are cases when one of the primers poorly matches to the SINE sequences in a given species and that is a reason why no 55 bp (AB)-PCR-product is observed in those cases (Borodulina and Kramerov, 1999). Although in the present work we have applied the (A)-PCR method only to those species in which the 55 bp (AB)-PCR-products were synthesized, one can use this approach in species whose genomic DNAs are unable to initiate synthesis of such short PCR-products. Given the lack of the short (AB)-PCR product is due to the poor matching of SINEs to primer B (whereas primer A matches well), (A)-PCR can yield a positive result. Also, in the case of poor matching of primer A (but not B), primer B can be used instead of primer A. Such a reaction can be named (B)-PCR. Finally, it is possible (and preferable) to carry out PCR with a primer not complementary but identical to the B box (this reaction will amplify DNA fragments resembling (A)-PCR-products but flanked with B box sequences). DNA fragments amplified by one of the (B)-PCR variants can be cloned and their terminal regions sequenced. As a result, the part of a novel SINE sequence will be identified. Subsequent inter-SINE-PCR of genomic DNA with a primer specific for this SINE will allow the determination of the full consensus sequence of the SINEs family studied.

4.2. Simple and complex SINEs

Using the (A)-PCR we isolated two novel SINEs families: Das-1 from the genome of nine-banded armadillo and Rhin-1 from the genome of great horseshoe bat. SINEs of edentate have not been studied previously, and Das-1 is the first example of SINEs from a representative of this mammalian order. Rhin-1, which occurs only in families Rhinolophidae and Hipposideridea, is the second SINE known in bats. Interestingly, the previously described SINE VES1 (Borodulina and Kramerov, 1999; Kawai et al., 2002) is present in bats from families Vespertilionidae, Mollosidae, and Emballonuridae, but not Rhinolophidae and Hipposideridea. The character of Rhin-1 and VES1 distribution among bats (Fig. 5C) supports the existence of the two phylogenetic clusters in Chiroptera (Teeling et al., 2002).

Das-1 and Rhin-1 structures are quite different. Das-1 is very short (90 bp) and has simple structure, whereas Rhin-1 is significantly longer (200–270 bp) and includes two monomers, one of which seems unrelated to tRNA, and its copies often contain long additional sequences. The length of most SINEs known is 150–170 bp. Perhaps such length is optimal with respect to SINE retroposition capacity. SINEs of several families are known to be longer and more complex. Such SINEs often consist of two monomers, e.g. Alu (Deininger et al., 1981), galago type II SINE (Daniels and Deininger, 1983), MEN (Serdobova and Kramerov, 1998), B1-dID (Kramerov and Vassetzky, 2001), and B4/RSINE (Lee et al., 1998). Rhin-1 is additionally characterized by internal insertions and duplications that allow it to be considered as complex SINEs. Perhaps the increase in size and structure complexity is an alternative strategy for effective SINEs.

Das-1 is probably the shortest and simplest SINE known: it consists only of a tRNA-derived region and a short A-rich tail. There are two other examples of simple SINEs: ID, from rodents (Kim et al., 1994) and Vic-1, from vicunas and camels (Lin et al., 2001). It should be noted that ID (subtypes 2, 3, and 4 but not 1) and Vic-1 have long A-rich tails containing a few other than A nucleotides at conserved positions (Fig. 3B). These areas are rather conserved, therefore the 5' part of the A-rich tails of ID and Vic-1 could be considered as a conserved tRNA-unrelated region. Thus, Das-1 is the most prominent member of this group, which we propose to term "simple SINEs". There is one more SINE consisted of only a tRNA-derived region and a poly(A) tail: CYN from flying lemur (Schmitz and Zischler, 2003). However, the vast majority of its sequenced copies (37 out of 38) are rather long because they consist of two or three monomers (Piskurek et al., 2003; Schmitz and Zischler, 2003).

Interestingly, CYN is related to isoleucine tRNA, whereas all three other families of simple SINEs are derived from alanine tRNA. Judging from their taxonomic distribution, Das-1, Vic-1, and ID originated from alanine tRNA independently of one another. ID occurs only in rodents. Vic-1 is characteristic only of family Camelidea, but not of the other artiodactyls (Lin et al., 2001). Das-1 seems to be inherent in only armadillos but not in other edentate families, such as sloths and anteaters. We have not investigated the taxonomic distribution of Das-1. However, Lin et al. (2001) studied the distribution of Vic-1 among mammals and observed no hybridization signal for sloth and anteater DNAs. Considering the high level of sequence similarity between Vic-1 and Das-1, this result suggests the lack of $tRNA^{Ala}$-derived SINEs, such as Vic-1 or Das-1, in the sloth and anteater genomes. One can speculate that alanine tRNA has some structural feature, which allows it to generate very simply organized but effectively propagated

SINEs. Perhaps this tRNA has high affinity to one or more components of the retroposition machinery.

The analysis of Das-1 and Vic-1 consensus sequences shows that both these SINEs families preserved the high level of similarity to $tRNA^{Ala}$ in the course of evolution (Fig 3B). In the consensus sequences, there are no insertions and deletions, except for the 5′end. It is of interest that Das-1 and Vic-1 (with possibly independent origin) have similar changes relative from $tRNA^{Ala}$ in their 3′region (indicated by arrows in Fig. 3B). Such similarity can be an example of the convergence at the molecular level. In future, analysis of the nucleotide substitutions that occurred in the tRNA sequence during the evolution of Das-1 and Vic-1 or other simple SINEs, can shed light on the significance of particular substitutions for SINEs activity.

Acknowledgements

We are grateful to Drs V. Matveev, A. Borisenko, R. DeBry, and V. Grechko for providing tissues or DNAs. This study was supported by the Russian Foundation for Basic Research (02-04-49772) and by Ministry of Science, Industry, and Technology of Russian Federation (43.073.1.1.2508).

References

Bannikova, A.A., Matveev, V.A., Kramerov, D.A., 2002. Using inter-SINE-PCR to study mammalian phylogeny. Genetika (Mosc.) 38, 853–864.

Borodulina, O.R., Kramerov, D.A., 1999. Wide distribution of short interspersed elements among eukaryotic genomes. FEBS Lett. 457, 409–413.

Borodulina, O.R., Kramerov, D.A., 2001. Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. Mamm. Genome 12, 779–786.

Daniels, G.R., Deininger, P.L., 1983. A second major class of Alu family repeated DNA sequences in a primate genome. Nucleic Acids Res. 11, 7595–7610.

Deininger, P.L., Batzer, M.A., 2002. Mammalian retroelements. Genome Res. 12, 1455–1465.

Deininger, P.L., Jolly, D.J., Rubin, C.M., Friedmann, T., Schmid, C.W., 1981. Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J. Mol. Biol. 151, 17–33.

Eickbush, T.H., Furano, A.V., 2002. Fruit flies and humans respond differently to retrotransposons. Curr. Opin. Genet. Dev. 12, 669–674.

Jurka, J., Zietkiewicz, E., Labuda, D., 1995. Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the Mesozoic era. Nucleic Acids Res. 23, 170–175.

Kass, D.H., Batzer, M.A., 1995. Inter-Alu polymerase chain reaction: advancements and applications. Anal. Biochem. 228, 185–193.

Kawai, K., et al., 2002. Intra- and interfamily relationships of Vespertilionidae inferred by various molecular markers including SINE insertion data. J. Mol. Evol. 55, 284–301.

Kim, J., Martignetti, J.A., Shen, M.R., Brosius, J., Deininger, P., 1994. Rodent BC1 RNA gene as a master gene for ID element amplification. Proc. Natl. Acad. Sci. U. S. A. 91, 3607–3611.

Kramerov, D.A., Vassetzky, N.S., 2001. Structure and origin of a novel dimeric retroposon B1-diD. J. Mol. Evol. 52, 137–143.

Kramerov, D., Vassetzky, N., Serdobova, I., 1999. The evolutionary position of dormice (Gliridae) in Rodentia determined by a novel short retroposon. Mol. Biol. Evol. 16, 715–717.

Lee, I.Y., et al., 1998. Complete genomic sequence and analysis of the prion protein gene region from three mammalian species. Genome Res. 8, 1022–1037.

Lin, Z., Nomura, O., Hayashi, T., Wada, Y., Yasue, H., 2001. Characterization of a SINE species from vicuna and its distribution in animal species including the family Camelidae. Mamm. Genome 12, 305–308.

Nikaido, M., Nishihara, H., Hukumoto, Y., Okada, N., 2003. Ancient SINEs from African endemic mammals. Mol. Biol. Evol. 20, 522–527.

Okada, N., 1991. SINEs. Curr. Opin. Genet. Dev. 1, 498–504.

Okada, N., Ohshima, K., 1995. Evolution of tRNA-derived SINEs. In: Maraia, R.J. (Ed.), The Impact of Short Interspersed Elements (SINEs) on the Host Genome. R.G. Landes Company, Austin, pp. 61–79.

Piskurek, O., Nikaido, M., Boeadi, Baba, M., Okada, N., 2003. Unique mammalian tRNA-derived repetitive elements in dermopterans: the t-SINE family and its retrotransposition through multiple sources. Mol. Biol. Evol. 20, 1659–1668.

Redston, M.S., Kern, S.E., 1994. Klenow co-sequencing: a method for eliminating “stops”. BioTechniques 17, 286–288.

Roos, C., Schmitz, J., Zischler, H., 2004. Primate jumping genes elucidate strepsirrhine phylogeny. Proc. Natl. Acad. Sci. U. S. A. 101, 10650–10654.

Russo, T., et al., 1986. Structure and in vitro transcription of tRNA gene clusters containing the primers of MuLV reverse transcriptase. Eur. J. Biochem. 158, 437–442.

Russo, T., Duilio, A., Ammendola, R., Costanzo, F., Cimino, F., 1987. Nucleotide sequence of a mouse tRNA gene cluster. Nucleic Acids Res. 15, 8562.

Schmitz, J., Zischler, H., 2003. A novel family of tRNA-derived SINEs in the colugo and two new retrotransposable markers separating dermopterans from primates. Mol. Phylogenet. Evol. 28, 341–349.

Schmitz, J., Ohme, M., Suryobroto, B., Zischler, H., 2002. The colugo (Cynocephalus variegatus, Dermoptera): the primates’ gliding sister? Mol. Biol. Evol. 19, 2308–2312.

Schmitz, J., Churakov, G., Zischler, H., Brosius, J., 2004. A novel class of mammalian-specific tailless retropseudogenes. Genome Res. 14, 1911–1915.

Serdobova, I.M., Kramerov, D.A., 1998. Short retroposons of the B2 superfamily: evolution and application for the study of rodent phylogeny. J. Mol. Evol. 46, 202–214.

Shedlock, A.M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. BioEssays 22, 148–160.

Shimamura, M., et al., 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388, 666–670.

Stoneking, M., et al., 1991. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res. 7, 1061–1071.

Teeling, E.C., Madsen, O., Van den Bussche, R.A., de Jong, W.W., Stanhope, M.J., Springer, M.S., 2002. Microbat paraphyly and the convergent evolution of a key innovation in Old World rhinolophoid microbats. Proc. Natl. Acad. Sci. U. S. A. 99, 1431–1436.

Vassetzky, N.S., Ten, O.A., Kramerov, D.A., 2003. B1 and related SINEs in mammalian genomes. Gene 319, 149–160.