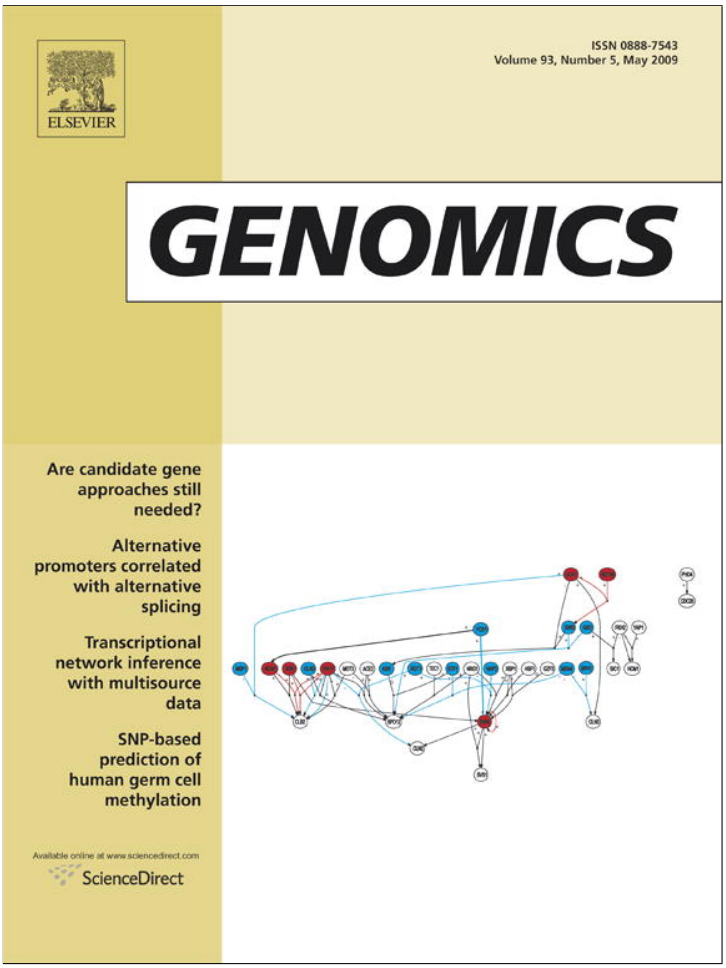

Genomics 93 (2009) 494–500

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

5S rRNA-derived and tRNA-derived SINEs in fruit bats

Konstantin P. Gogolevsky, Nikita S. Vassetzky, Dmitri A. Kramerov *

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov St., Moscow 119991, Russia

ARTICLE INFO

Article history:
Received 3 December 2008
Accepted 4 February 2009
Available online 12 February 2009

Keywords:
SINEs
LINEs
Fruit bats
5S rRNA
Evolution

ABSTRACT

Most short retroposons (SINEs) descend from cellular tRNA of 7SL RNA. Here, four new SINEs were found in megabats (Megachiroptera) but neither in microbats nor in other mammals. Two of them, MEG-RS and MEG-RL, descend from another cellular RNA, 5S rRNA; one (MEG-T2) is a tRNA-derived SINE; and MEG-TR is a hybrid tRNA/5S rRNA SINE. Insertion locus analysis suggests that these SINEs were active in the recent fruit bat evolution. Analysis of MEG-RS and MEG-RL in comparison with other few 5S rRNA-derived SINEs demonstrates that the internal RNA polymerase III promoter is their most invariant region, while the secondary structure is more variable. The mechanisms underlying the modular structure of these and other SINEs as well as their variation are discussed. The scenario of evolution of MEG SINEs is proposed.

Introduction

Repetitive elements constitute the bulk (or at least a significant fraction) of the genome in higher eukaryotes. They are divided into tandem and interspersed repeats. Some interspersed repeats called DNA transposons translocate via a DNA intermediate, while the replication of retrotransposons, which amount to 30–40% of the mammalian genome, involves the transcription of their genomic copies followed by reverse transcription of an RNA intermediate and DNA integration into the genome. Retrotransposons are divided into LTR-containing elements, long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs).

SINEs are short (100–500 bp) elements that encode no proteins; they are transcribed by cellular RNA polymerase III, while the subsequent reverse transcription and integration into the genome is mediated by the LINE machinery [1–3]. Unlike all other transposons, most SINE families do not have a common origin but emerged independently in different host lineages from cellular and LINE modules [4]. The 5′ end of most SINEs is derived from a cellular tRNA; however, in some SINEs it can also descend from 7SL RNA (Alu, B1, and their derivatives) [5–7] or 5S rRNA (SINE3, some DeuSINEs, and Ped-1) [8–10]. The origin of the central part of most SINE families is obscure, although some otherwise unrelated SINEs share the same 'core' region (three such cores are known to date [9,11,12]). The 3′ end of many but not all SINEs descends from the 3′ end of the partner LINE [13], the retropositional machinery of which is utilized.

tRNA-derived SINEs are most universal and exist in all animal and plant taxa, where SINEs were identified. In particular, we have previously described two tRNA-derived SINE families in bats (Chiroptera), VES in Yangochiroptera and Rhin-1 in Rhinolophidae and Hipposideridea [14,15]. At the same time, 7SL RNA-derived SINEs are limited to mammals (to be precise, to the Supraprimate clade). 5S rRNA-derived SINEs were found in fishes (SINE3) but were likely active in the common ancestor of vertebrates [8,9]. Recently, we have found a novel active 5S rRNA-derived SINE Ped-1 in rodent springhare [10]. The distribution of Ped-1 was clearly restricted to a single rodent family; it was found in neither other rodent families nor other vertebrates. Here, we tried to identify other 5S rRNA-derived SINEs in mammalian genomes. As a result, two more such SINEs have been found in fruit bats (or megabats, Megachiroptera) as well as a tRNA-derived SINE and a hybrid tRNA/5S rRNA SINE.

Materials and methods

DNA samples

This study focused on megabats, and we used DNA from four genera covering major megabat lineages. For reference, we used eight species from five microbat families, as well as several representative of four mammalian orders. Tissue samples of bats were kindly provided by A. Borisenko (Zoological Museum of Moscow State University) and V. Matveev (Moscow State University). DNA was isolated from fresh, frozen, or ethanol-preserved tissues (liver, placenta, or muscle) by incubation with proteinase K followed by phenol/chloroform extraction. DNA was quantified by fluorometry using Hoechst dye 33258.

Dot-blot hybridization and PCR

Three approaches were used to detect SINEs of interest in the genomes, dot hybridization, PCR, and computer analysis (see below).

* Corresponding author. Fax: +7 495 135 1405.
E-mail address: kramerov@eimb.ru (D.A. Kramerov).

0888-7543/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2009.02.001

In addition, PCR was used to prepare hybridization probes. Genomic DNAs (500 ng) were incubated in 10 μl of 0.5 M NaOH for 1 h at 37 °C. After the incubation, 20 volumes of 6× SSC, 6% formaldehyde, and 0.025 M NaH_2PO_4 were added, and the DNA was transferred to a Hybond N membrane. Hybridization and washing conditions were the same as in colony screening of the genomic libraries. Filters were autoradiographed and the hybridization signal intensity measured with a Cyclone phosphorimager (Packard).

PCR reaction mixtures (25 μl) contained 1 U Taq polymerase, 200 μM dNTP, reaction buffer, 0.1 ng megabat genomic DNA and a pair of specific primers (Table S1). Thirty cycles of 95 °C for 1 min, annealing temperature (Table S1) for 1 min, and 72 °C for 1 min were performed to amplify SINE fragments. PCR products were separated by electrophoresis in 2% or 4% agarose gel and stained with ethidium bromide. For preparation of hybridization probes, DNA fragments were eluted from gel and labeled by 20 cycles of PCR with α[^{32}P] dATP [14].

Library construction and screening

We used our common strategy to isolate SINEs from the genomes. Genomic DNA fragments of the desired size range were used to construct a library, which was screened by hybridization with 5S rRNA probe, and positive clones were analyzed by sequencing. Genomic DNA (5.0 μg) was digested with EcoRI and HindIII and separated by electrophoresis in 1% agarose gel. DNA fragments of 0.5–1.2 kb were collected by reverse electrophoresis on a DEAE membrane inserted in the gel. DNA was eluted from the membrane in 400 μl of 1 M NaCl, 1 mM EDTA, and 10 mM Tris–HCl, pH 8.0, for 30 min at 60 °C. The isolated genomic fragments (0.2 μg) were ligated into 0.1 μg pGEM3Z, digested with EcoRI and HindIII, and used to transform XL-1 Blue *Escherichia coli* cells. 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 5S rRNA probe. Nitrocellulose filters were washed in 0.1× SSC and 0.1% SDS at 42 °C, and positive colonies were identified by autoradiography. Colonies with hybridization signal were selected and purified by two additional rounds of colony hybridization. The number of positive clones during library screening was used to estimate the number of MEG-RL and MEG-RS copies.

DNA sequencing and computer analysis

The clones isolated by library screening were sequenced. Sequence databases of megabats and some other mammals were screened for MEG SINEs. The sequences obtained by both approaches were analyzed using the standard computer methods. Cloned SINE-containing DNA fragments were sequenced using standard M13 primers, the BigDye Terminator sequencing kit, and an ABI Prism 3100-Avant sequencer (Applied Biosystems). The nucleotide sequences of cloned DNA fragments of Egyptian fruit bat (*Rousettus aegyptiacus*) and cave nectar bat (*Eonycteris spelaea*) were deposited in GenBank under accession numbers EU853261–EU853287.

MEG SINE sequences were also extracted from the whole-genome shotgun bank of large flying fox (*Pteropus vampyrus*) and from all megabat sequences available in GenBank using the FASTA program. Multiple alignments were produced by the ClustalW program and manually adjusted in GeneDoc. All consensus sequences were corrected for hypervariable CpG/CpA/TpG methylation targets. Consensus sequences of MEG SINEs were generated from sequences obtained in this work and found in sequence banks. The number of MEG SINE copies in the *P. vampyrus* genome was roughly estimated from the number of SINE sequences n found in WGS sequences of total length L using equation $N = \frac{n \times 3 \times 10^9}{L}$ (3×10^9 states for the length of haploid mammalian genome). RNA secondary structure was predicted for 5S rRNA-derived sequences using the mFold 3.2 web server with default parameters [16]. The best prediction was used.

Results

The bulk of currently active mammalian SINEs are mobilized by LINE1 (L1); however, two more LINEs can also be involved: L2 [17] and Bov-B [18]. Recently, we have described two new SINEs in springhare (Rodentia) also mobilized by Bov-B [10]. One of them (Ped-2) is tRNA-derived and another (Ped-1) is 5S rRNA-derived. We tried to evaluate the distribution of Ped-1 by dot hybridization of 5S rRNA probe with the genomic DNA of a wide range of vertebrates. Weak positive signals were observed in many species, which likely corresponded to a considerable number of 5S rRNA genes and pseudogenes in their genomes (e.g., ~2000 in human, mouse, and rat). Not surprisingly, strong hybridization signals were observed in springhare but also in the common frog, green iguana, and Egyptian fruit bat. While such signals in amphibians and, likely, reptiles can be attributed to much higher number of 5S rRNA (pseudo)genes (at least 20 times that in human), these data suggested a 5S rRNA-derived SINE in the fruit bat [10].

MEG SINEs in fruit bats

We tested this suggestion by screening the genomic libraries of Egyptian fruit bat (*Rousettus aegyptiacus*, Megachiroptera) and cave nectar bat *Eonycteris spelaea* with 5S rRNA probe. Some positive clones were sequenced and proved to include a typical SINE (MEG-RL). It included a 5′ part similar to 5S rRNA and a short unrelated part.

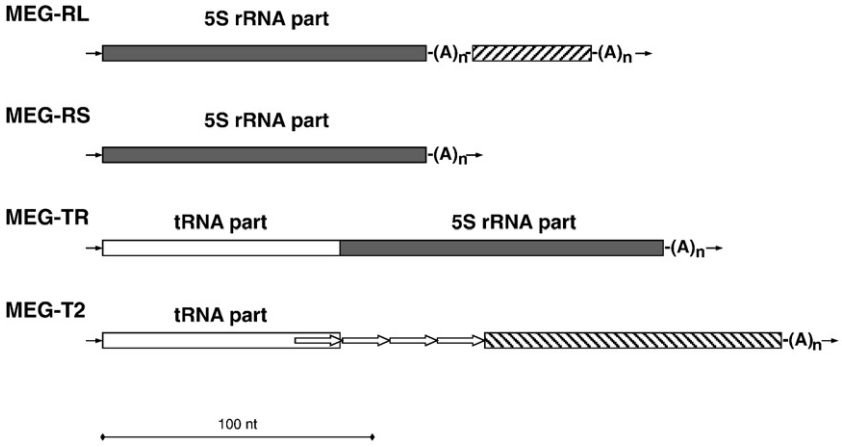

Fig. 1. Diagram of fruit bat SINEs. The 5S rRNA- and tRNA-related regions are grayed and empty, respectively. Arrows indicate direct repeats. For other explanations, see text.

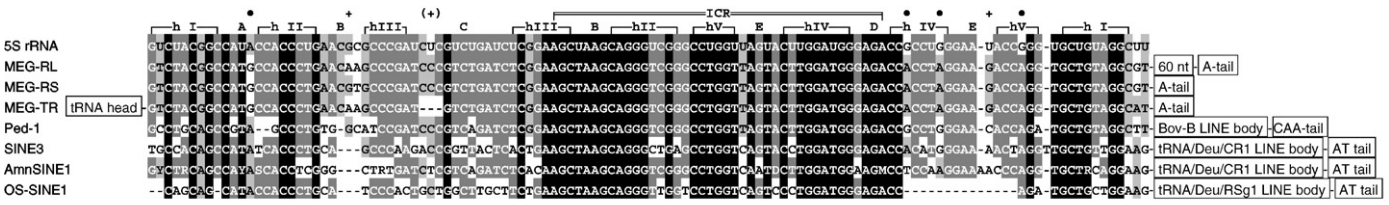

Fig. 2. Multiple alignment of consensus sequences of 5S rRNA-derived SINEs. Human 5S rRNA with the internal RNA polymerase III promoter (ICR), helices (hI–hV) and loops (A–E) is shown above (all available mammalian 5S rRNA sequences are identical; [44]). Other than 5S rRNA-derived blocks are shown as boxes. MEG-xx, fruit bat SINEs; Ped-1, SINE from rodent springhare [10]; SINE3, fish SINE [8]; AmnSINE1, SINE3 element of amniotes (human) [9]; OS-SINE1, salmon SINE [9]. Positions differing between 5S rRNA and all MEG SINEs or MEG-RL + MEG-TR are marked by '●' or '+', respectively. (We considered differences with nearly all individual sequences significant, which explains why not all differences with consensus sequences are marked.)

In addition, there were many shorter sequences (MEG-RS) that we initially identified as 5S rRNA pseudogenes and a single element with a tRNA-derived 5′-part and 5S rRNA-derived 3′-part (MEG-TR). Searching similar sequences yielded these SINEs in the genome of large flying fox (*Pteropus vampyrus*) as well as another SINE with the same tRNA part and unrelated 3′-region (MEG-T2). Finally, four SINEs were found in the genomes of fruit bats (Fig. 1): 5S rRNA-derived *MEG-RL* and *MEG-RS*, hybrid tRNA/5S rRNA-derived *MEG-TR*, and tRNA-derived *MEG-T2*.

MEG-RL SINE

MEG-RL consensus is 213 bp long (Fig. S1). It includes a 120-bp 5′-region related to 5S rRNA (90% identity to mouse 120-nt 5S rRNA) followed by a short A-rich spacer (from 9 to 31 bp; typically $A_6TA_7TA_2$), a 60-bp 3′-region of unknown origin, and an A-rich tail (either pure poly(A) or $(A_{3-6}B)_n$, where B is C, G, or T). Nearly all MEG-RL copies are flanked by short direct repeats or target sequence duplications (TSDs) presumably generated during retroposition. The MEG-RL sequences are very similar to each other (their median similarity is 81%, which is high for a SINE family).

MEG-RS SINE

MEG-RS consensus is 135 bp long and includes only a 5S rRNA-related region (90% identity to mouse 120-bp 5S rRNA) and an A-rich tail (pure poly(A) or AT-rich). Most MEG-RS copies have TSDs and their median similarity is also 81% (Fig. S2). MEG-RS is nearly identical to MEG-RL (just two mismatches apart from the tail) but shorter.

Overall, the structure of MEG-RS is so simple that it can be considered a 5S rRNA pseudogene. Moreover, it clearly originated from a 5S rRNA pseudogene. However, we prefer to call it a SINE taking into account two considerations. First, the number of MEG-RS copies in the genome is significantly higher than the number of 5S rRNA pseudogenes in most other mammals including microbats. Second, and more important, is that the majority of MEG-RS copies have positions different from 5S rRNA sequence (marked by '●' in Fig. 2). This indicates that most MEG-RS copies descended from a modified 5S rRNA pseudogene, a master copy, rather than from 5S rRNA gene, and that this master copy provided better templates for amplification in the genome than 5S rRNA gene. For comparison, we extracted around 200 sequences from the human genomic bank with similarity throughout the 5S rRNA sequence region. Contrary to MEG-RS, there were no major differences from the 5S rRNA sequence, and their consensus sequence perfectly matched that of 5S rRNA. The same pattern was observed for 5S rRNA pseudogenes in the dog genome (data not shown). Moreover, the median similarity of human and dog 5S rRNA pseudogenes was by ~20% lower than that of MEG-RS indicating more recent origin of MEG-RS.

MEG-TR SINE

This composite 220-bp-long SINE includes two regions of cellular origin. At the 5′-region, a short (6-bp) G-rich tag (typical of most tRNA-derived SINEs) is followed by a $tRNA^{Val}$-related region (96% identity to mouse $tRNA^{Val}_{UAC}$ over a 73-nt region). A 9-bp spacer (not poly(A)) links it to a 5S rRNA-derived region (88% identity to mouse 120-nt 5S rRNA). The tail is AT-rich; most commonly, several (TA_{2-3}) repeats are followed by a pure A stretch of different length. TSD is found in nearly all MEG-TR copies (Fig. S3). The 5S rRNA-derived

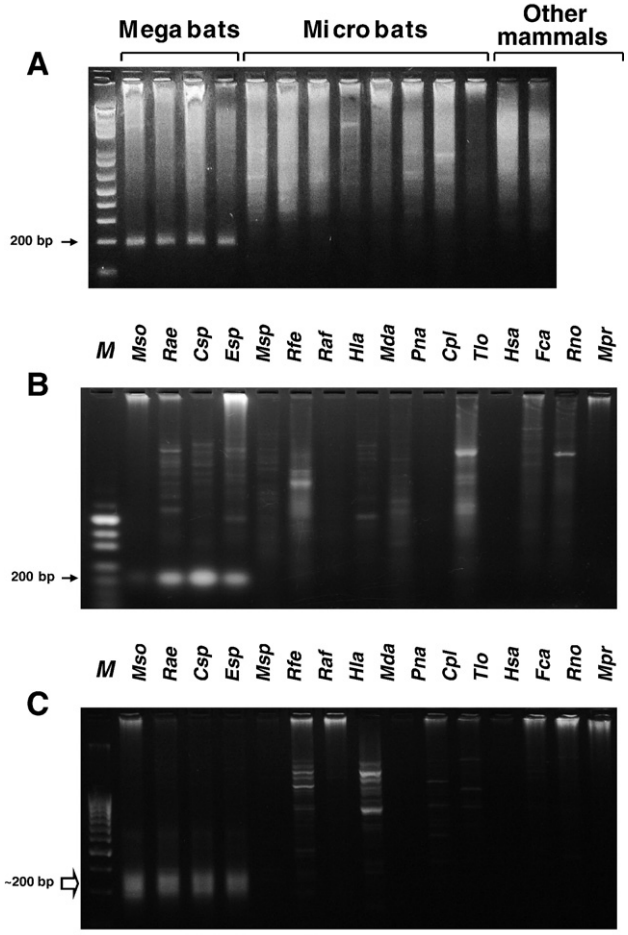

Fig. 3. PCR of genomic DNA from bats and other mammals with primers specific for MEG-RL (A), MEG-TR (B), and MEG-T2 (C). Megabats: *Mso*, *Macroglossus sobrinus* (long-tongued fruit bat); *Rae*, *Rousettus aegyptiacus* (Egyptian fruit bat); *Esp*, *Eonycteris spelaea* (cave nectar bat); *Csp*, *Cynopterus sphinx* (greater short-nosed fruit bat). Microbats: *Tlo*, *Taphozous longimanus* (long-winged tomb bat); *Msp*, *Megaderma spasma* (lesser false vampire bat); *Raf*, *Rhinolophus affinis* (intermediate horseshoe bat); *Rfe*, *Rhinolophus ferrumequinum* (Greater horseshoe bat); *Hla*, *Hipposideros larvatus* (intermediate roundleaf bat); *Cpl*, *Chaerephon plicata* (wrinkle-lipped free-tailed bat); *Mda*, *Myotis daubentoni* (Daubenton's bat); *Pna*, *Pipistrellus nathusii* (Nathusius' pipistrelle). Other mammals: *Hsa*, human (Homo sapiens); *Fca*, *Felis catus* (domestic cat); *Rno*, *Rattus norvegicus* (Norway rat); *Mpr*, *Macroscelides proboscideus* (short-eared elephant shrew); M, size marker. PCR mixture contained 0.1 ng (megabats) or 10 ng genomic DNA (other species).

region is similar to those in MEG-RL and MEG-RS (95 and 93% identity, respectively). The median similarity of MEG-TR copies is 85%.

MEG-T2 SINE

MEG-T2 is a tRNA-derived SINE with the 5′-structure similar to that of MEG-TR (94% identity over a 83-nt region): a short G-tag followed by a $tRNA^{Val}$-related region (92% identity to mouse tRNA over a 73-nt region). It is followed by a region of variable length composed of several (four to twelve in the copies available) tandem repeats CNCCRGG (which are sometimes organized into longer ~dimers **CCCCRGG**tg**CGCCAGG**) (Fig. S4). The body part is not related to any known sequences. The tail is AT-rich (in many cases, relatively long stretches of A and T). Nearly all MEG-T2 copies are flanked by TSDs. The median similarity of MEG-T2 copies (excluding the variable region with tandem repeats) is 74%.

Distribution and abundance of Meg SINEs

Many 5S rRNA-derived SINEs (both MEG-RL and MEG-RS) and a single MEG-TR copy in Egyptian fruit bat were cloned and sequenced. One MEG-RL was found among bank sequences of another species of this genus, Leschenault's rousette (*R. leschenaultia*). Several MEG-RL and MEG-RS sequences were obtained from the genome of cave nectar bat. Finally, many copies of all four MEG SINEs (28–74 each) were found among the bank sequences of large flying fox (*Pteropus vampyrus*).

The distribution of MEG SINEs was further studied using genomic PCR. The primers specific for MEG-RL revealed the expected 200-bp PCR product in four tested megabats (Pteropodidae) but neither in microbats (Microchiroptera) nor in other tested mammals (Fig. 3A). Likewise, the expected bands were generated by PCR with primers specific for MEG-TR (Fig. 3B) and MEG-T2 (Fig. 3C) in megabats but not in other bats or mammals. Note the fuzzy band for MEG-T2 SINE (Fig. 3C) due to the variable tandem repeat region.

Similarly, dot hybridization of genomic DNAs with probes specific for 5S rRNA, the 3′ part of MEG-RL, and the tRNA region of MEG-TR and -T2 SINEs shows strong signals only for megabat species (Fig. 4). While signals corresponding to 5S rRNA pseudogenes are visible in other mammals (although not as intense as in fruit bats) (Fig. 4A), these data altogether demonstrate that MEG SINEs are limited to megabats.

Our rough estimation of the number of MEG SINE copies in the genomes based on their frequency in the large flying fox sequence bank yielded that there are $\sim 10^4$ MEG-RL, 7×10^3 MEG-RS, 5×10^3 MEG-TR, and 4×10^3 MEG-T2. Similar estimates were obtained for MEG-RL and MEG-RS in Egyptian fruit bat based on the number of positive clones in the genomic library 3×10^3 and 5×10^3, respectively.

Since MEG SINEs with flanking sequences were available for four species (including actively sequenced *P. vampyrus*), we analyzed the specific sites of integration. We identified such sites in *P. vampyrus* bank sequences in all but one case (22 out of 23) (data not shown). The SINE (namely, MEG-RS) was missing in only one of these 22 sites (MEG-RS was present in cave nectar bat and absent in large flying fox), thus, indicating a MEG SINE insertion after the genera *Eonycteris* and *Pteropus* separated.

5S rRNA-derived SINEs

MEG SINEs described here in fruit bats and rodent Ped-1 SINE described in our previous publication [10] substantially increase the number of 5S rRNA-derived SINEs known to date. Fig. 2 shows multiple alignment of their 5S rRNA-derived region (a single SINE3 variant was included except for AmnSINE1, also a SINE3 variant, since it has notable differences in the 5S rRNA region). One can see that the most invariant region in these SINEs is the central region precisely corresponding to the internal RNA polymerase III promoter ('internal control region' or ICR in Fig. 2), while both 5′- and 3′-regions are more variable.

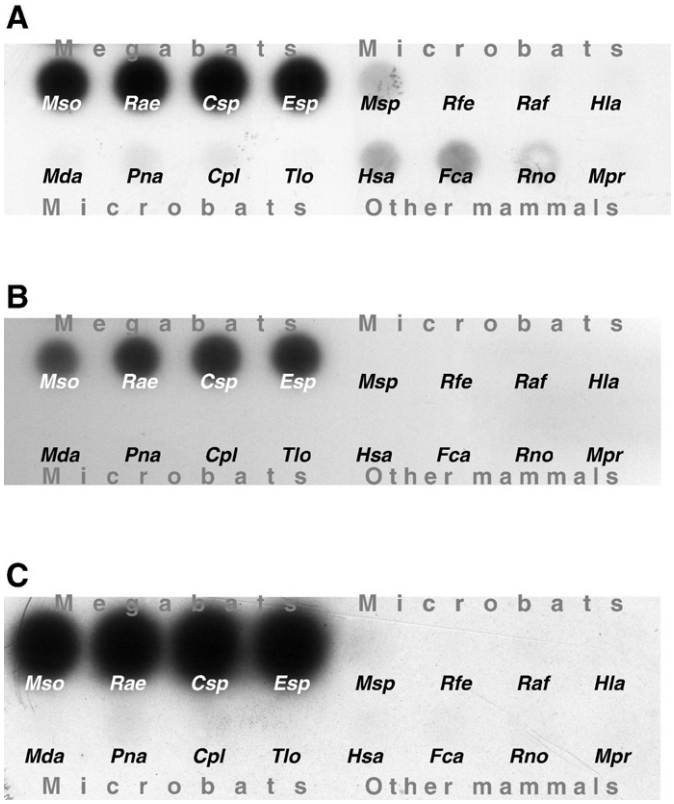

Fig. 4. Dot hybridization of genomic DNA from bats and other mammals to probes specific for 5S rRNA (A), 3′-portion (beyond the 5S rRNA-derived region) of MEG-RL (B), and 5′-portion (tRNA-derived) of MEG-TR/MEG-T2 (C). The probe specific for the central part of MEG-T2 yielded the same hybridization pattern as the 5′-tRNA probe (data not shown). For species designation, see Fig. 3.

We tried to analyze the secondary structure of 5S rRNA-derived SINEs using the mFold energy-minimization algorithm [16] and consensus sequences as the best approximate to the master copies. This approach has recently proved consistent with experimentally determined secondary structures for at least two SINEs [19]. One can recognize the 5S rRNA secondary structure in MEG SINEs and in Ped-1 also resembles that in 5S rRNA (and the free energies are similar), while it is not the case in other SINEs (and the free energy is higher). Note that the sequence identity with 5S rRNA also decreases from MEG SINEs (71–74%) and Ped-1 (67%) to SINE3, AmnSINE, and OS-SINE1 (60, 57, and 49%, respectively) (Fig. 5).

Discussion

Unlike other types of mobile elements, SINEs are transcribed by RNA polymerase III. Most SINEs descend from cellular tRNAs, although the first discovered SINEs, human Alu and mouse B1, originate from 7SL RNA. While some cellular RNA genes transcribed by RNA polymerase III (U6 and 7SK) utilize promoter elements located outside of the transcribed sequence, 7SL and transfer RNA genes have internal promoter elements (within the transcribed sequence), and this is essential for mobile elements. This logic prompted Weiner [20] to predict a new class of SINEs derived from 5S rRNA, another abundant cellular RNA with an internal promoter transcribed by RNA polymerase III. Almost immediately such element (SINE3) has been found in zebrafish [8] and later in other fish species and, in minor quantities, even in birds and mammals (AmnSINE) [9]. A similar OS-SINE was found in salmonid fishes. These SINEs have a complex structure: 5′ – 5S rRNA-derived region – tRNA-derived region – Deu-domain (a core region shared by several sometimes otherwise unrelated SINEs) – LINE-derived region – AT-rich tail – 3′. The

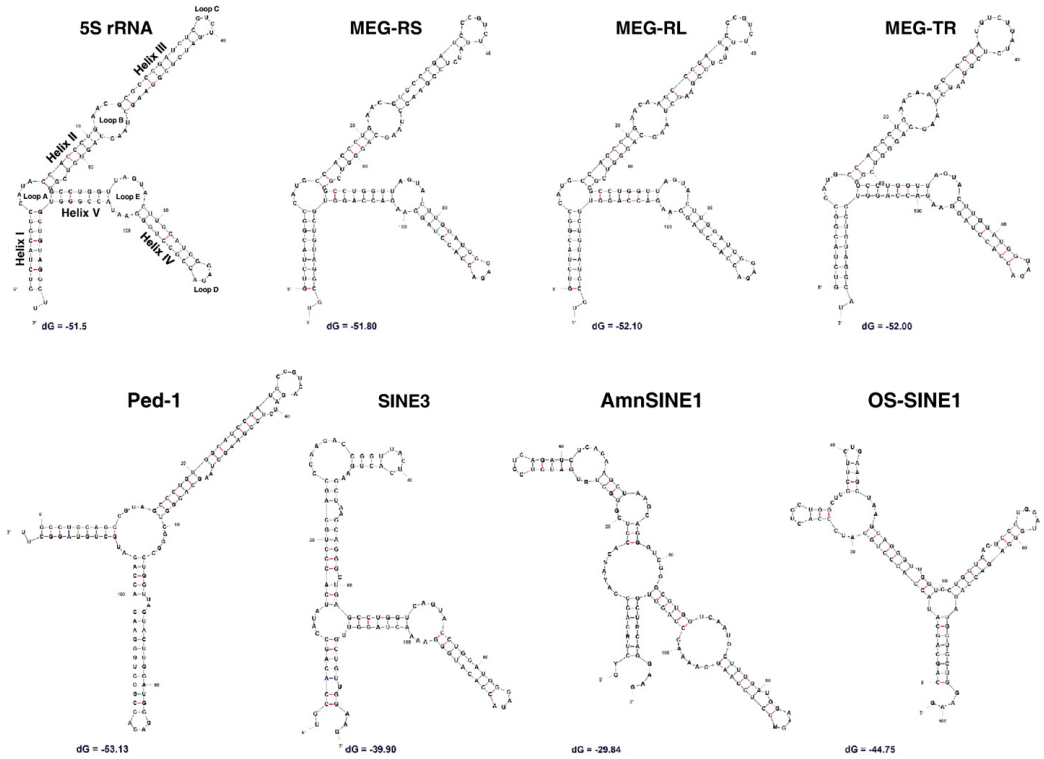

Fig. 5. Predicted secondary structures of 5S rRNA-derived region of SINE consensus sequences (mFold). Since the predicted structure was the same for the full-length and 5S rRNA-derived region of all SINEs except AmnSINE1, the predicted secondary structures of only the 5S rRNA-derived region are given (to make free energy values comparable). Helices I–V and loops A–E [45] are shown in human 5S rRNA. MEG-*xx*, fruit bat SINEs; Ped-1, SINE from rodent springhare [10]; SINE3, fish SINE [8]; AmnSINE1, SINE3 element of amniotes (human) [9]; OS-SINE1, salmon SINE [9].

SINE3/AmnSINE family essentially differs from OS-SINE1 only in the ~70-bp LINE-derived region originating from CR1 and RSg-1 LINEs, respectively [9].

Recently, we have described one more 5S rRNA-derived SINE, Ped-1, in the genome of springhare, the only species representing rodent family Pedetidae. The structure of this Ped-1 is not so complex, 5′ – 5S rRNA-derived region – LINE-derived region – AT-rich tail – 3′; although they have an unusual bipartite LINE-derived region (originating from the 5′-region and extreme 3′-region of LINE) typical of SINEs mobilized by Bov-B LINE [10].

Here we describe three more 5S rRNA-derived SINEs in fruit bats. Their structure is overall simpler compared to other 5S rRNA-derived SINEs. Apart from the AT-tail, MEG-RL includes a 5′-part derived from 5S rRNA and a 60-bp 3′-region of unknown origin (Fig. 1). MEG-RS is even simpler, it lacks the 60-nt region. MEG-TR is a dimeric SINE with a tRNA-derived sequence at the 5′ end and a 5S rRNA-derived sequence at the 3′ end. The structure of MEG-T2 is most complex of all fruit bat SINEs, it has a tRNA-derived region at the 5′ end, a 3′-region of unknown origin, and a tandem repeat region between them.

These SINEs considerably extend the number of patterns of 5S rRNA-derived SINEs known to date. MEG-RS is the first instance of a *simple* SINE [15] derived from 5S rRNA rather than from tRNA (thus, not a *t*-SINE; see [21,22]). Unlike other tRNA/5S rRNA hybrid SINEs, MEG-TR contains a full-length tRNA-derived region and has a tRNA rather than a 5S rRNA at the 5′ end (where its transcription is initiated). Finally, MEG-T2 has an internal tandem repeat region of variable length.

Alignment of consensus sequences of all 5S rRNA-derived SINE families (Fig. 2) shows that the internal control region (ICR) that constitutes the RNA polymerase III promoter is the most invariant. Moreover, the three elements of the promoter (boxes A, IE, and C) can also be recognized as the most similar sites within the ICR. There is one more invariant sequence element in the 3′ end of all 5S rRNA SINEs.

According to mFold predictions, the secondary structure of 5S rRNA was conserved in the MEG SINEs and Ped-1, the structure of other SINEs deviated from the original structure. Note that the primary and secondary structure similarities with 5S rRNA correlated. Overall, the conservation of the original secondary structure seems not always essential. It makes sense considering that the 5S rRNA helices and loops are required for the interaction with ribosomal proteins, and there is no data supporting that such interaction can be beneficial for SINEs. At the same time, SINE transcription requires the RNA polymerase III promoter, and ICR remains invariant in all 5S rRNA-derived SINEs. Clearly, only further studies can address this issue.

Parsimony considerations and the sequence similarity within each MEG SINE family suggest the following scenario of their origin. MEG-T2 was the first SINE in fruit bats as indicated by the higher sequence variation of MEG-T2 copies vs. 5S rRNA-derived MEG SINEs (median similarity of 74 and 81–85%, respectively). We failed to find any simpler tRNA-derived SINEs (particularly, tRNA pseudogene-like simple SINE) in megabats indicating that the MEG-T2 precursors were not efficient retroposons. Independently, a 5S rRNA pseudogene with some modifications (indicated by '●' in Fig. 2) has become a successful MEG-RS SINE. MEG-RS was likely the first 5S rRNA-derived SINE in megabats; apart from parsimony considerations, it is confirmed by the appearance of extra modifications in MEG-RL and MEG-TR (indicated by '+' in Fig. 2). Finally, MEG-TR SINE appeared as a hybrid of the tRNA and 5S rRNA lineages of MEG SINEs, and its recency is supported by the relative homogeneity of MEG-TR sequences.

The capacity of SINEs to combine into dimers is not limited to 5S rRNA-derived SINEs. It is well known for 7SL RNA-derived SINEs that tend to dimerize with tRNA-derived SINEs. Many such examples are known in mammals, both tRNA–7SL RNA (numerous ID–B1 SINEs in rodents [23–25], SINE type II in primates [26], and Tu types I and II in treeshrews [27]) and 7SL RNA–tRNA (numerous B1–ID SINEs in rodents [25,28]). Homodimers are less common (only one 7SL–7SL SINE is known, Alu in primates [29]) but not exclusive (tRNA–tRNA SINEs include CYN in flying lemurs [30,31], DAS-II in xenarthrans [32], Twin in mosquitoes [33], and some SB SINEs in mustard plants [34]). There are also tRNA–tRNA SINEs that are not homodimeric as their parts descend from different tRNA species (IDl-Geo in rodents [35] and CELE45 in nematodes (unpublished data)).

On the contrary, here we report the first monomeric 5S rRNA-derived SINEs; nevertheless the trend of 5S rRNA-derived SINEs to dimerization is apparent. Thus, one can expect 7SL RNA–5S rRNA (or 5S rRNA–7SL RNA) hybrid SINEs to be described as sequences of more species become available, while the discovery of homodimeric 5S rRNA-derived SINE seems less probable (but not impossible).

MEG-T2 has an interesting region of variable length composed of short tandem repeats. Such regions are not uncommon in the 3′-part of SINEs. For instance, SOR was described in shrews as a SINE with three internal tandem repeats of about 20 bp [36]. However, database sequences of the common shrew (*Sorex araneus*) also contain SOR variants with one, two, four, and even five such repeats (data not shown). A region of variable length composed of shorter repeats $(CT)_n$ is found in CAN SINE of carnivores [37], rabbit C repeat [38], DIP [24], and Tu III in tree shrews [39]. While the DNA-mediated mechanisms similar to those of microsatellite variation are possible, the ability of LINE reverse transcriptase to jump on short direct repeats (see [4,40]) seems a more probable mechanism of such variation in SINEs.

Many SINEs have their 3′ portion similar to that of the partner LINE; however, most mammalian SINEs have no LINE-related sequences and their tail is A-rich (or AT-rich). Such SINEs are mobilized by LINE1 (L1). L1 is active in most mammals, and it was active in the bat ancestor [41]. Overall, L1 is the most probable partner LINE of MEG SINEs.

The presence of MEG SINEs at more than 20 integration sites in the *Rousettus*/*Eonycteris* and *Pteropus* genera representing distinct fruit bat clades suggests that MEG SINEs were active in the megabat ancestor. A single integration event detected after the divergence of *Eonycteris* and *Pteropus* suggests a recent decline in MEG SINE activity. This decline agrees with the recent data suggesting that L1 is currently inactive in megabats [41]. At the same time, if MEG SINEs were indeed mobilized by L1, the MEG-RS integration event after the separation of *Eonycteris* and *Pteropus* indicates that L1 remained active in the early evolution of megabats. Although it seems less probable, this case can be attributed to unfixed population dynamics or precise elimination of this SINE copy.

Analysis of the MEG SINE structures and of many other SINEs suggests that they are designed using a 'construction set' including standard 'pieces' such as cellular RNAs transcribed by RNA polymerase III or 3′ end of a partner LINE. Other pieces look like random sequences since we know nothing about their function or origin, although some of them, e.g., several 'cores' found in otherwise unrelated SINEs must have some (yet unidentified) function.

Why different SINE structures arise? First, a new combination of pieces can prove an effective retrotransposon competitive with the previous design(s). Second, SINEs depend on LINEs, and a modification of LINE or the appearance of a new LINE can underlie using a different 'piece' fitting the preferences of that LINE. Apart from random mechanisms joining the 'pieces,' there may be specific mechanisms of SINE assembly. First, it is template switch common for reverse transcriptases (in particular, in LINEs [2,42,43]), which can combine different parts of LINE, cellular, and SINE RNAs. Second, it can be integration of reverse transcriptase transcripts into the genomic copies of other retroposons (SINEs or LINEs) or cellular pseudogenes (e.g., 5S rRNA). Third, it can be recombination between different SINEs, LINEs, or cellular pseudogenes sharing similar sequence regions. As more different SINE types are described and analyzed,

these and, possibly, other mechanisms of their origin and evolution will become approved.

Acknowledgments

This work was supported by the Russian Foundation for Basic Research (project nos. 07-04-00462-a and 08-04-00350-a) and the Molecular and Cellular Biology Program of the Russian Academy of Sciences.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2009.02.001.

References

[1] J. Jurka, Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 1872–1877.
[2] M. Kajikawa, N. Okada, LINEs mobilize SINEs in the eel through a shared 3′ sequence, Cell 111 (2002) 433–444.
[3] M. Dewannieux, C. Esnault, T. Heidmann, LINE-mediated retrotransposition of marked Alu sequences, Nat. Genet. 35 (2003) 41–48.
[4] D.A. Kramerov, N.S. Vassetzky, Short retroposons in eukaryotic genomes, Int. Rev. Cytol. 247 (2005) 165–221.
[5] E. Ullu, C. Tschudi, Alu sequences are processed 7SL RNA genes, Nature 312 (1984) 171–172.
[6] Y. Quentin, Emergence of master sequences in families of retroposons derived from 7SL RNA, Genetica 93 (1994) 203–215.
[7] J.O. Kriegs, G. Churakov, J. Jurka, J. Brosius, J. Schmitz, Evolutionary history of 7SL RNA-derived SINEs in Supraprimates, Trends Genet. 23 (2007) 158–161.
[8] V.V. Kapitonov, J. Jurka, A novel class of SINE elements derived from 5S rRNA, Mol. Biol. Evol. 20 (2003) 694–702.
[9] H. Nishihara, A.F. Smit, N. Okada, Functional noncoding sequences derived from SINEs in the mammalian genome, Genome Res. 16 (2006) 864–874.
[10] K.P. Gogolevsky, N.S. Vassetzky, D.A. Kramerov, Bov-B-mobilized SINEs in vertebrate genomes, Gene 407 (2008) 75–85.
[11] N. Gilbert, D. Labuda, CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 2869–2874.
[12] I. Ogiwara, M. Miya, K. Ohshima, N. Okada, V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit, Genome Res. 12 (2002) 316–324.
[13] K. Ohshima, N. Okada, SINEs and LINEs: symbionts of eukaryotic genomes with a common tail, Cytogenet. Genome Res. 110 (2005) 475–490.
[14] O.R. Borodulina, D.A. Kramerov, Wide distribution of short interspersed elements among eukaryotic genomes, FEBS Lett. 457 (1999) 409–413.
[15] O.R. Borodulina, D.A. Kramerov, PCR-based approach to SINE isolation: simple and complex SINEs, Gene 349 (2005) 197–205.
[16] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, Nucleic Acids Res. 31 (2003) 3406–3415.
[17] N. Gilbert, D. Labuda, Evolutionary inventions and continuity of CORE-SINEs in mammals, J. Mol. Biol. 298 (2000) 365–377.
[18] N. Okada, M. Hamada, The 3′ ends of tRNA-derived SINEs originated from the 3′ ends of LINEs: A new example from the bovine genome, J. Mol. Evol. 44 (1997) 52–56.
[19] F.J. Sun, S. Fleurdepine, C. Bousquet-Antonelli, G. Caetano-Anolles, J.M. Deragon, Common evolutionary trends for SINE RNA structures, Trends Genet. 23 (2007) 26–33.
[20] A.M. Weiner, SINEs and LINEs: the art of biting the hand that feeds you, Curr. Opin. Cell Biol. 14 (2002) 343–350.
[21] O. Piskurek, N. Okada, Simple and complex SINEs: a brief critical comment, Gene 375 (2006) 110.
[22] O.R. Borodulina, D.A. Kramerov, t-SINE or simple SINE? Can be both, Gene 375 (2006) 111.
[23] I.Y. Lee, et al., Complete genomic sequence and analysis of the prion protein gene region from three mammalian species, Genome Res. 8 (1998) 1022–1037.
[24] I.M. Serdobova, D.A. Kramerov, Short retroposons of the B2 superfamily: evolution and application for the study of rodent phylogeny, J. Mol. Evol. 46 (1998) 202–214.
[25] N.A. Veniaminova, N.S. Vassetzky, D.A. Kramerov, B1 SINEs in different rodent families, Genomics 89 (2007) 678–686.
[26] G.R. Daniels, P.L. Deininger, A second major class of Alu family repeated DNA sequences in a primate genome, Nucleic Acids Res. 11 (1983) 7595–7610.
[27] H. Nishihara, Y. Terai, N. Okada, Characterization of novel Alu- and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins, Mol. Biol. Evol. 19 (2002) 1964–1972.
[28] D.A. Kramerov, N.S. Vassetzky, Structure and origin of a novel dimeric retroposon B1-dID, J. Mol. Evol. 52 (2001) 137–143.
[29] P.L. Deininger, D.J. Jolly, C.M. Rubin, T. Friedmann, C.W. Schmid, Base sequence studies of 300 nucleotide renatured repeated human DNA clones, J. Mol. Biol. 151 (1981) 17–33.
[30] O. Piskurek, M. Nikaido, Boeadi, M. Baba, N. Okada, Unique mammalian tRNA-derived repetitive elements in dermopterans: the t-SINE family and its retrotransposition through multiple sources, Mol. Biol. Evol. 20 (2003) 1659–1668.
[31] J. Schmitz, H. Zischler, A novel family of tRNA-derived SINEs in the colugo and two new retrotransposable markers separating dermopterans from primates, Mol. Phylogenet. Evol. 28 (2003) 341–349.
[32] G. Churakov, A.F. Smit, J. Brosius, J. Schmitz, A novel abundant family of retroposed elements (DAS-SINEs) in the nine-banded armadillo (*Dasypus novemcinctus*), Mol. Biol. Evol. 22 (2005) 886–893.
[33] C. Feschotte, N. Fourrier, I. Desmons, C. Mouches, Birth of a retroposon: the Twin SINE family from the vector mosquito *Culex pipiens* may have originated from a dimeric tRNA precursor, Mol. Biol. Evol. 18 (2001) 74–84.
[34] J.M. Deragon, X. Zhang, Short interspersed elements (SINEs) in plants: origin, classification, and use as phylogenetic markers, Syst. Biol. 55 (2006) 949–956.
[35] K.P. Gogolevsky, D.A. Kramerov, Short interspersed elements (SINEs) of the Geomyoidea superfamily rodents, Gene 373 (2006) 67–74.
[36] O.R. Borodulina, D.A. Kramerov, Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure, Mamm. Genome 12 (2001) 779–786.
[37] N.S. Vassetzky, D.A. Kramerov, CAN—a pan-carnivore SINE family, Mamm. Genome 13 (2002) 50–57.
[38] R.C. Hardison, R. Printz, Variability within the rabbit C repeats and sequences shared with other SINES, Nucleic Acids Res. 13 (1985) 1073–1088.
[39] O.A. Ten, O.R. Borodulina, N.S. Vasetskii, N.I. Oparina, D.A. Kramerov, Repetitive sequences of the tree shrew genome (Mammalia, Scandentia), Mol. Biol. 40 (2006) 63–71.
[40] N.S. Vassetzky, O.A. Ten, D.A. Kramerov, B1 and related SINEs in mammalian genomes, Gene 319 (2003) 149–160.
[41] M.A. Cantrell, L. Scott, C.J. Brown, A.R. Martinez, H.A. Wichman, Loss of LINE-1 activity in the megabats, Genetics 178 (2008) 393–404.
[42] A. Bibillo, T.H. Eickbush, The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates, J. Mol. Biol. 316 (2002) 459–473.
[43] A. Buzdin, et al., The human genome contains many types of chimeric retrogenes generated through in vivo RNA recombination, Nucleic Acids Res. 31 (2003) 4385–4390.
[44] M. Szymanski, M.Z. Barciszewska, V.A. Erdmann, J. Barciszewski, 5S ribosomal RNA database, Nucleic Acids Res. 30 (2002) 176–178.
[45] M. Szymanski, M.Z. Barciszewska, V.A. Erdmann, J. Barciszewski, 5 S rRNA: structure and interactions, Biochem. J. 371 (2003) 641–651.