Short interspersed elements (SINEs) of the *Geomyoidea* superfamily rodents

Konstantin P. Gogolevsky, Dmitri A. Kramerov *

Laboratory of Eukaryotic Genome Evolution, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov Street, Moscow 119991, Russia

Received 23 November 2005; received in revised form 4 January 2006; accepted 9 January 2006

Available online 6 March 2006

Received by C.W. Schmid

**Abstract**

A new short interspersed element (SINE) was isolated from the genome of desert kangaroo rat (*Dipodomys deserti*) using single-primer PCR. This SINE consists of two monomers: the left monomer (IDL) resembles rodent ID element and other tRNAAla(CGC)-derived SINEs, whereas the right one (Geo) shows no similarity with known SINE sequences. PCR and hybridization analyses demonstrated that IDL-Geo SINE is restricted to the rodent superfamily Geomyoidea (families Geomyidea and Heteromyidea). Isolation and analysis of IDL-Geo from California pocket mouse (*Chaetodipus californicus*) and Botta’s pocket gopher (*Thomomys bottae*) revealed some species-specific features of this SINE family. The structure and evolution of known dimeric SINEs are discussed.

© 2006 Elsevier B.V. All rights reserved.

**Keywords:** Repetitive DNA; Dimeric short retroposon; Geomyoidea; Heteromyidea; Genome; PCR

1. Introduction

Short interspersed elements (SINEs), also known as short retroposons, are 80–400 nucleotides (nt) long repetitive DNA sequences that propagate in eukaryotic genomes via transcription followed by reverse transcription (Okada, 1991; Deininger and Batzer, 2002; Kramerov and Vassetzky, 2005). Usually one or several SINE families can be found in a eukaryotic genome. Each SINE family is represented by $10^3$–$10^5$ individual copies in the genome. Their sequences could vary significantly (5–35%) depending on the time when these copies emerged. About 70 SINE families have been isolated from the genomes of mammals, reptiles, fishes, mollusks, ascidia, insects, and flowering plants. SINEs are transcribed by RNA polymerase III (pol III) due to the presence of the internal promoter for pol III in their 5′ region composed of two boxes (A and B) separated by a 30–40nt sequence.

Two originally discovered SINE families, B1 in rodents (Krayev et al., 1980) and Alu in primates (Deininger et al., 1981), are thought to originate from 7SL RNA, a component of the signal recognition particles (SRPs) that are involved in translation of secreted proteins (Ullu and Tschudi, 1984). With the only exception of 5S rRNA-derived SINEs in zebrafish (Kapitonov and Jurka, 2003), all other SINE families are descendants of tRNA molecules. Usually the 5′ region of these genetic elements demonstrates a reasonable sequence similarity to a particular tRNA, and is followed by a unique sequence characteristic of each SINE family. The 3′ end of SINEs is represented either by an A-rich tail (in mammals and plants) or by short tandem repeats (in fish, reptiles, and invertebrates). In addition to B1 SINE, which is characteristic of all 30 rodent families (Vassetzky et al., 2003), several other, tRNA-derived, SINE families have been found in the genomes of this mammalian order. The first described tRNA-derived short retroposon is B2 (Krayev et al., 1982). This 190-nt long SINE can be found in four rodent families: Muridae (mice and rats), Cricitidae (hamsters, voles, and gerbils), Spalacidae (mole rats), and Rhizomyidae (bamboo rats) (Serdobova and Kramerov, 1998). The next identified SINE, ID, (∼100nt) contains a tRNAAla-derived sequence and an A-rich tail (Milner et al., 1984; Kim et al., 1994). ID is presumably present in the genomes of all rodents, however, the number of copies is usually not high ($10^2$–$10^3$ per genome), except for rats and deer mice ($1.5 \times 10^5$ and $2.5 \times 10^4$, respectively) (Kass et al., 1996). ID originates from small cellular RNA, BC1, found in the nervous system of primates.
tissue (Kim et al., 1994), whereas alanine tRNA^{GCG} is the evolutionary precursor of BC1 RNA. However, dimeric SINEs composed of ID and B1 moiety can be very abundant, for instance, B4 from rat and mouse (Lee et al., 1998; Waterston et al., 2002) or B1-dID from squirrels (Sciuridae) and dormice (Gliridae) (Kramerov et al., 1999; Kramerov and Vassetzky, 2001). These SINEs have a different order of the monomers, ID–B1 in B4 and vice versa in B1-dID.

One more dimeric SINE family, MEN, has been isolated from palm squirrel (Menetus berdmorei). This SINE is characteristic of the tribe Callosciurini, but is absent from other squirrels (Serdobova and Kramerov, 1998). The left monomer of MEN resembles ID but has an additional sequence at the 3′ end, whereas the right one is B1. Finally, a monomeric 200nt SINE referred to as DIP is typical of two rodent families: jerboas (Dipodidae) and birch mice (Zapodidae) (Serdobova and Kramerov, 1998).

Two major points make the isolation and characterization of novel SINE families of great importance for molecular genetics and evolution. First, it helps us to elucidate the patterns of the structural organization, functioning, and evolution of these genetic elements. Second, SINE families or their individual genomic copies are perfect phylogenetic markers (Shimamura et al., 1997; Stoneking et al., 1997; Serdobova and Kramerov, 1998; Kramerov et al., 1999; Shedlock and Okada, 2000; Nikaido et al., 2003; Roos et al., 2004).

Recently, Borodulina and Kramerov (2005) proposed a method for isolation of novel SINE families based on PCR of genomic DNA with a single primer specific to box A of the pol III promoter. Here we further develop this approach. The first 20 nucleotides of the B2 superfamily SINES demonstrate high degree of conservation; in particular, B2 and DIP display 100% identity in this region (Serdobova and Kramerov, 1998). Accordingly, we carried out PCR using a primer specific to the 5′ ends of the B2 and DIP to study the genomes of rodent families Heteromyidae and Geomyidae. This method allows novel SINEs to be identified using very small quantities of genomic DNA. With as little as 200ng of the kangaroo rat DNA, we managed to identify a novel SINE family, IDL-Geo. This SINE has been found only in heteromyids (pocket mice) and geomyids (pocket gophers), i.e., within the North American superfamily Geomyoidea.

2. Materials and methods

2.1. Specimens

DNAs from Dipodomys deserti (desert kangaroo rat), Chaetodipus californicus (California pocket mouse), Liomys pictus (painted spiny pocket mouse), Perognatus parvus (great basin pocket mouse), Thomomys bottae (Botta’s pocket gopher) were kindly provided by Dr. R. DeBry (University of Cincinnati,}

---

Fig. 1. Alignment of D. deserti PCR-products. Seven upper sequences were obtained using primer 1 ( ), and five lower sequences using primer 2 ( ). The top sequence represents the deduced consensus of D. deserti SINE. The nucleotide sequences of primer 1 and 2 are shown above the consensus. GenBank accession numbers: AY588979–AY588990.
USA). T. bottae DNA was donated by Dr. F. Catzeffis (Institut des Sciences de l’Évolution, Universite Montpellier II, France). Pedetes capensis DNA was a generous gift of Dr. T. Robinson (University of Stellenbosch, Matieland, South Africa). The sources of all other samples were described elsewhere (Borodulina and Kramerov, 2001; Gogolevskaya and Kramerov, 2002).

2.2. PCR methods

Single-primer PCR was carried out in 50μl mix containing 10 ng of genomic DNA, 0.3 μM primer 1 or 2 (Fig. 1), 200μM dNTP, 1U Taq-polymerase and reaction buffer (67mM Tris–HCl pH 8.6, 2.5mM MgCl2, 16.6mM (NH4)2SO4 and 0.001% HCl pH 8.6, 2.5mM MgCl2, 16.6mM (NH4)2SO4 and 0.001% Triton X-100) and included 30 amplification cycles (95 °C, 1 min; 56 °C, 1 min; 72 °C, 1 min). After electrophoretic separation in agarose gel, PCR products longer than 300 bp were collected by reverse electrophoresis on DEAE membrane (NA 45, Schleicher and Schuell) placed into the loading well.

IDL-Geo SINEs were amplified from the genomes of various rodents using 27 PCR cycles (95 °C, 1 min; 56 °C, 1 min; 72 °C, 1 min) with 0.3 or 10ng genomic DNAs and primers 1 and 2. PCR products were separated by electrophoresis in gel containing 3% Nusieve agarose and 1% regular agarose. For radioactive labeling, 2% of isolated PCR product was added to PCR mix (25μl) containing 30 μCi [α-32P]dATP, and 20 PCR cycles (95 °C, 1 min; 56 °C, 1 min; 72 °C, 20s) were carried out. For library construction, the single-primer PCR product purified by agarose electrophoresis was cloned into pGEM-T (Promega) according to the manufacturer’s instructions. Colony hybridization was carried out at 60 °C in 4× SSC, 6% formaldehyde, and 0.025M NaH2PO4 for 1h at 37 °C. After the incubation, 20 volumes of 0.5M NaOH for 1h at 37 °C. After the incubation, 20 volumes of 0.5M NaOH were added, and DNA was applied to a Hybond N membrane using a dot-blot apparatus. Hybridization and washing conditions as recommended by the manufacturer. The rodent species included in this study

<table>
<thead>
<tr>
<th>Family</th>
<th>Code*</th>
<th>Scientific name</th>
<th>Common name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muridae</td>
<td>Mmu</td>
<td>Mus musculus</td>
<td>House mouse</td>
</tr>
<tr>
<td>Cricetidae</td>
<td>Pda</td>
<td>Perognatus parvus</td>
<td>Great basin pocket mouse</td>
</tr>
<tr>
<td>Spalacidae</td>
<td>Smi</td>
<td>Dipodomys deserti</td>
<td>Painted spiny pocket mouse</td>
</tr>
<tr>
<td>Rhizomyidae</td>
<td>Tsp</td>
<td>Eremodipus lichtensteini</td>
<td>Liechtenstein’s jerboa</td>
</tr>
<tr>
<td>Ctenodactylidae</td>
<td>Cgu</td>
<td>Ctenodactylus gundi</td>
<td>Gundy</td>
</tr>
<tr>
<td>Caviidae</td>
<td>Cpo</td>
<td>Cavia porcellus</td>
<td>Guinea pig</td>
</tr>
<tr>
<td>Hydrochaeridae</td>
<td>Hca</td>
<td>Hydrochoerus capybara</td>
<td>Capybara</td>
</tr>
<tr>
<td>Dasyproctidae</td>
<td>Mac</td>
<td>Myoprocta acouchy</td>
<td>Green acouchi</td>
</tr>
<tr>
<td>Pedetidae</td>
<td>Pca</td>
<td>Pedetes capensis</td>
<td>Springhare</td>
</tr>
<tr>
<td>Heteromyidae</td>
<td>Cca</td>
<td>Chaetodipus californicus</td>
<td>California pocket mouse</td>
</tr>
<tr>
<td>Dasyproctidae</td>
<td>Dde</td>
<td>Dipodomys deserti</td>
<td>Desert kangaroo rat</td>
</tr>
<tr>
<td>Leporidae</td>
<td>Lpi</td>
<td>Lepus americanus</td>
<td>Painted spiny pocket mouse</td>
</tr>
<tr>
<td>Geomyidae</td>
<td>Ppa</td>
<td>Perognathus parvus</td>
<td>Great basin pocket mouse</td>
</tr>
</tbody>
</table>

2.3. DNA sequencing and computer analysis

Cloned PCR products were sequenced using the dideoxy method with Sequenase 2.0 (Amersham Bioscience) and standard (forward and reverse) M13 primers according to the manufacturer’s instructions with modifications (Redston and Kern, 1994). Full-length IDL-Geo copies were sequenced using the BigDye Terminator sequencing kit and an ABI Prism 3100-Avant sequencer (Applied Biosystems). The nucleotide sequences were aligned using the Clustal V method and the GeneDoc alignment editor (Nicholas and Nicholas, 1997).

2.4. Dot-blot hybridization

Rodent genomic DNAs (50 or 500 ng) were incubated in 10μl of 0.5M NaOH for 1h at 37 °C. After the incubation, 20 volumes of 6× SSC, 6% formaldehyde, and 0.025M NaH2PO4 were added, and DNA was applied to a Hybond N membrane using a dot-blot apparatus. Hybridization and washing conditions as recommended by the manufacturer. The rodent species included in this study

<table>
<thead>
<tr>
<th>Family</th>
<th>Code*</th>
<th>Scientific name</th>
<th>Common name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muridae</td>
<td>Mmu</td>
<td>Mus musculus</td>
<td>House mouse</td>
</tr>
<tr>
<td>Cricetidae</td>
<td>Pda</td>
<td>Perognatus parvus</td>
<td>Great basin pocket mouse</td>
</tr>
<tr>
<td>Spalacidae</td>
<td>Smi</td>
<td>Dipodomys deserti</td>
<td>Painted spiny pocket mouse</td>
</tr>
<tr>
<td>Rhizomyidae</td>
<td>Tsp</td>
<td>Eremodipus lichtensteini</td>
<td>Liechtenstein’s jerboa</td>
</tr>
<tr>
<td>Ctenodactylidae</td>
<td>Cgu</td>
<td>Ctenodactylus gundi</td>
<td>Gundy</td>
</tr>
<tr>
<td>Caviidae</td>
<td>Cpo</td>
<td>Cavia porcellus</td>
<td>Guinea pig</td>
</tr>
<tr>
<td>Hydrochaeridae</td>
<td>Hca</td>
<td>Hydrochoerus capybara</td>
<td>Capybara</td>
</tr>
<tr>
<td>Dasyproctidae</td>
<td>Mac</td>
<td>Myoprocta acouchy</td>
<td>Green acouchi</td>
</tr>
<tr>
<td>Pedetidae</td>
<td>Pca</td>
<td>Pedetes capensis</td>
<td>Springhare</td>
</tr>
<tr>
<td>Heteromyidae</td>
<td>Cca</td>
<td>Chaetodipus californicus</td>
<td>California pocket mouse</td>
</tr>
<tr>
<td>Dasyproctidae</td>
<td>Dde</td>
<td>Dipodomys deserti</td>
<td>Desert kangaroo rat</td>
</tr>
<tr>
<td>Leporidae</td>
<td>Lpi</td>
<td>Lepus americanus</td>
<td>Painted spiny pocket mouse</td>
</tr>
<tr>
<td>Geomyidae</td>
<td>Ppa</td>
<td>Perognathus parvus</td>
<td>Great basin pocket mouse</td>
</tr>
</tbody>
</table>

* Three-letter abbreviations for scientific names of the species used in PCR and dot-hybridization analyses (Fig. 3).  

b The species analyzed by dot-hybridization only.
well as radioactive probe preparation were the same as in colony hybridization screening of the genome libraries.

3. Results

3.1. PCR amplification of SINE-containing DNA fragments

Since SINEs have not been previously studied in the genomes of rodent families Heteromyidea and Geomyidea, the main goal of the present study was to search for SINEs of the B2 superfamily in these rodents. Low quantities of the available heteromyid and geomyid DNA substantiated the use of our PCR-based method of SINE identification (Borodulina and Kramerov, 2005). PCR was performed using a single primer specific to the 5′ region of mouse B2 and the jerboa DIP elements (primer 1) and 10 ng of genomic DNA from kangaroo rat (Heteromyidea) or Botta’s pocket gopher (Geomyidea). The PCR products were visible in agarose gel as a smear in a length range from 200 to 3000 nt (data not shown). This result suggested that the kangaroo rat and pocket gopher genomes contain B2-related SINEs.

3.2. SINE cloning and sequencing

PCR products from the kangaroo rat genome longer than 300 bp were eluted from agarose gel and cloned into plasmid vector. The resulting library was screened by hybridization with the same 32P-labeled PCR product. Plasmid DNA was isolated from the positive clones and sequenced from both ends (Fig. 1, seven top sequences).

3.3. Sequence alignment and comparison

In five clones (Dde1, Dde3, Dde11, Dde28, and Dde40), the primer 1 sequence was immediately followed by about 200-nt regions with a high degree of similarity. All of them had A-rich regions at their 3′ terminus. The presence of sequences resembling A and B boxes of pol III promoter and of an A-rich tail suggested that these five clones represented SINE copies. Two more clones, Dde32 and Dde39, contained additional unique sequences between the primer 1 site and the sequences similar to those in five above-mentioned clones (due to occasional PCR priming in the sequences outside of SINE). As a result, two complete sequences of the D. deserti SINE were unexpectedly obtained.

3.4. Sequencing the 5′ terminal SINE region

In order to determine the 5′ SINE sequences, PCR was carried out using a primer complementary to the 3′ terminal region of the putative SINE (primer 2) and D. deserti genomic DNA. This reaction (called reverse PCR) yielded heterogeneous products that formed a smear after electrophoresis in agarose gel (data not shown). These products represent DNA fragments containing SINEs.

Fig. 3. Taxonomic distribution of the IDL-Geo SINE family. (A) PCR-based detection of the IDL-Geo in rodent genomes. Results of two experiments are shown. PCR-product of expected size (180 bp) is marked by an arrow. (B) Dot-hybridization of the labeled Geo-specific probe with genomic DNA from rodents. 50 ng genomic DNAs from geomyoids (Tbo, Lpi, Ppa, and Dde) and 500 ng DNAs from other species were blotted onto filter. See Table 1 for species names.

Fig. 4. Full-length nucleotide sequences of IDL-Geo SINEs from T. bottae, Botta’s pocket gopher (A) and C. californicus, California pocket mouse (B). Sequences whose names end in “i” and “g” were selected by screening of the libraries with ID- and Geo-specific probes, respectively. Sequences flanking the SINE are shown in lowercase; direct flanking repeats are underlined. GenBank accession numbers for C. californicus and T. bottae DNA sequences are DQ206796–DQ206803 and DQ206804–DQ206817, respectively. (C) Comparison of IDL-Geo consensus sequences from D. deserti (Dde), C. californicus (Cca), and T. bottae (Tbo). Direct repeats in variable region of D. deserti Geo-monomer are marked by arrows.
located between two neighboring SINE copies in the opposite orientation. After cloning and sequencing their terminal regions, we aligned these sequences with each other and with the previously obtained sequences (Fig. 1, five bottom sequences). This made it possible to deduce the full-length consensus sequence of the \( D. \) \( \text{deserti} \) SINE.

### 3.5. Structural features of the putative SINE and its relations to other known SINES

The length of different SINE copies varied from 172 to 202bp, and the mean sequence identity between the consensus and individual copies was 89%. The 5′ region (≈80bp) of the \( D. \) \( \text{deserti} \) SINE proved to be similar to the corresponding regions of rodent SINES of the B2 superfamily (Fig. 2). For instance, this region of the \( D. \) \( \text{deserti} \) SINE consensus sequence had 82%, 77%, and 83% similarity with the corresponding regions of ID, DIP, and MEN elements, respectively. In addition, \( D. \) \( \text{deserti} \) SINE demonstrated 74% sequence identity with the alanine tRNA\( \text{(CGC)} \), the most probable ancestor of the entire B2 superfamily. This 5′ region was followed by A-rich sequences. We assumed that the \( D. \) \( \text{deserti} \) SINE consisted of two parts (monomers) spaced by an A-rich linker similar to other dimeric SINES such as Alu, MEN, or B1-dID (Deininger et al., 1981; Serdobova and Kramerov, 1998; Kramerov and Vassetzky, 2001).

The similarity of the first (left) monomer to ID element coupled with the pan-rodent distribution of ID made it a good candidate ancestor of the first monomer of the \( D. \) \( \text{deserti} \) SINE; accordingly, we named it IDL (ID-Like). The second (right) monomer showed no similarity to any known SINES or tRNA, although the sequence similar to box A of pol III promoter was found in its 5′ region (Fig. 1). As shown below, the genomes of all analyzed rodents of the Geomyoidea superfamily (Heteromyidea and Geomyidea) carry sequences of this novel second monomer, which was named Geo.

### 3.6. Distribution of the novel SINE among different rodent families

The range of IDL-Geo SINE distribution among rodent families was evaluated by PCR with primers 1 and 2 and DNA of 16 rodent species belonging to 13 families (Table 1). In the first experiment, the same amount of genomic DNA (0.3ng) was added to the reaction mixtures for all species. PCR products of expected length (≈180bp) were observed in representatives of two families only: heteromyids (\( C. \) \( \text{californicus} \), \( D. \) \( \text{deserti} \), \( L. \) \( \text{pictus} \), \( P. \) \( \text{parvus} \)) and geomyids (\( T. \) \( \text{bottae} \)) (Fig. 3A). Interestingly, the amplified fragment of the slightly smaller size was observed in the case of gopher \( T. \) \( \text{bottae} \) (Fig. 3A, Tbo lane). A smaller PCR product (about 140nt) was also observed in beaver (Castoridae). Later, this fragment was sequenced but showed no similarity to the IDL-Geo SINE except for primer 1 and 2 sequences (data not shown). No PCR products of expected or similar length were found in other rodents tested (Fig. 3A), even when a 100-fold excess (10ng) of the genomic DNA was used (data not shown). These results suggest that IDL-Geo is characteristic of the superfamily Geomyoidea only. Dot-hybridization with Geo-specific probe further confirms that the distribution of this repetitive sequence is restricted to Geomyoidea (Fig. 3B).

In order to additionally confirm the presence of this SINE in rodent species other than \( D. \) \( \text{deserti} \), we cloned and sequenced the DNA fragments amplified by PCR in the described above experiment from the genomes of Great Basin pocket mouse \( P. \) \( \text{parvus} \) and Botta’s pocket gopher \( T. \) \( \text{bottae} \) (data not shown, GenBank accession numbers AY588991–AY588999 and AY589000–AY589008, respectively). The structure of the SINES in these two species proved similar to that in kangaroo rat. The \( P. \) \( \text{parvus} \) and \( T. \) \( \text{bottae} \) IDL-Geo copies showed 83% and 80% nucleotide identity with the \( D. \) \( \text{deserti} \) consensus sequence, respectively. Despite some sequence differences between the described elements in these three species, they undoubtedly belong to the same SINE family, which suggests that it had emerged before the divergence of these species.

### 3.7. IDL-Geo SINE from \( T. \) \( \text{bottae} \) and \( C. \) \( \text{californicus} \)

Considering the differences mentioned above and the complex structure of the IDL-Geo element, we verified the PCR data by sequencing several full genomic copies of the \( T. \) \( \text{bottae} \) SINE cloned from a traditional genomic DNA library. Such cloning became possible after larger amounts of \( T. \) \( \text{bottae} \) DNA became available to us. The genomic library was screened with both

### Table 2

<table>
<thead>
<tr>
<th>SINE</th>
<th>Left monomer (progenitor RNA)</th>
<th>Right monomer (progenitor RNA)</th>
<th>Taxonomic distribution</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDL-Geo</td>
<td>ID (tRNA( ^{\text{Ala}} ))</td>
<td>Geo (? )</td>
<td>Heteromyidae (pocket mice), Geomyidae (pocket gophers)</td>
<td>This article</td>
</tr>
<tr>
<td>B1-dID</td>
<td>B1 (7SL RNA)</td>
<td>ID (tRNA( ^{\text{Ala}} ))</td>
<td>Sciuridae (squirrels), Gliridae (dormice)</td>
<td>Kramerov and Vassetzky, 2001</td>
</tr>
<tr>
<td>B4 (R-SINE2)</td>
<td>ID (tRNA( ^{\text{Ala}} ))</td>
<td>B1 (7SL RNA)</td>
<td>Muridae (mice and rats)</td>
<td>Lee et al., 1998</td>
</tr>
<tr>
<td>MEN</td>
<td>(tRNA( ^{\text{Ala}} ))</td>
<td>B1 (7SL RNA)</td>
<td>( Menetes ) ( \text{berdmorei} ) (palm squirrel)</td>
<td>Serdobova and Kramerov, 1998</td>
</tr>
<tr>
<td>Au</td>
<td>L (7SL RNA)</td>
<td>R (7SL RNA)</td>
<td>Primates</td>
<td>Deininger et al., 1981</td>
</tr>
<tr>
<td>SINE type II</td>
<td>(tRNA( ^{\text{Ala}} ))</td>
<td>(7SL RNA)</td>
<td>Lorisiformes (Bush babies and lorises)</td>
<td>Daniels and Deininger, 1983;</td>
</tr>
<tr>
<td>CYN</td>
<td>(tRNA( ^{\text{Ile}} ))</td>
<td>(tRNA( ^{\text{Ile}} ))</td>
<td>Dermoptera (flying lemur)</td>
<td>Roos et al., 2004</td>
</tr>
<tr>
<td>Das-II</td>
<td>(tRNA( ^{\text{Ala}} ))</td>
<td>(tRNA( ^{\text{Ala}} ))</td>
<td>( Dasyus ) ( \text{novemcinctus} ) (armadillo)</td>
<td>Schmitz and Zischler, 2003;</td>
</tr>
<tr>
<td>Twin</td>
<td>(tRNA( ^{\text{Ala}} ))</td>
<td>(tRNA( ^{\text{Ala}} ))</td>
<td>( Culex ) ( \text{pipiens} ) (mosquito)</td>
<td>Piskurek et al., 2003; Churakov et al., 2005; Fesheotte et al., 2001</td>
</tr>
</tbody>
</table>
IDL- and Geo-specific probes. IDL-specific screening yielded eight full-length IDL-Geo and five monomeric ID elements, while six IDL-Geo copies and no Geo monomers were isolated using the Geo-specific probe (Fig. 4A). Similar to other SINEs, most IDL-Geo copies were flanked by short direct repeats. Interestingly, the Tb6 clone included two tandemly arranged IDL-Geo copies, with short direct repeats flanking the entire structure (IDL-Geo-IDL-Geo) that suggested the integration of this whole structure into the genome. Accordingly, this element can represent a tetrameric IDL-Geo subfamily.

Finally, the genomic library of heteromyid C. californicus was screened in a similar way. Three IDL-Geo (Fig. 4B) and two monomeric ID elements (not shown) were isolated by screening with ID probe. Geo-targeted screening resulted in five IDL-Geo copies (Fig. 4B) and a Geo monomer (not shown). Thus, in both C. californicus and T. bottae genomes, IDL-Geo occurred more frequently than monomeric ID and, especially, Geo. The number of IDL-Geo copies was estimated to be around $1 \times 10^5$.

Comparison of IDL-Geo consensus sequences from three geomyoid species (Fig. 4C) revealed their apparent similarity, although two variable regions were found. The first region is the A-rich linker between the IDL and Geo monomers. This linker has the lowest content of A residues in C. californicus. The second region is in the middle of the Geo monomer. It is the shortest in T. bottae, while, in D. deserti, it is a part of two 7-nt tandem repeats. This region is clearly A-rich in C. californicus, suggesting that the Geo monomer could be formed by the fusion of two short retroelements, and this region was the A-rich tail of the first of them. Cca-25 copy with $\text{(A)}_{17}$ run in this region (Fig. 4B) could represent an ancient variant of IDL-Geo. Alternatively, this might be a rapidly evolving region (particularly, due to DNA slippage).

4. Discussion

4.1. Single-primer PCR approach to the SINEs isolation

In the present study we used a novel approach for the identification and cloning of short dispersed repetitive sequences (SINEs). It is based on use of a single primer for amplification of SINEs from genomic DNA of species under study (single-primer PCR). This method should work best for cloning new SINEs given the structure of SINEs from related organisms has already been studied. Using this approach, we were able to clone SINE copies from D. deserti (Heteromyidea family) that has not been studied in this respect previously.

Several rodent SINEs display high sequence similarity in their 5′ ends (Serdobova and Kramerov, 1998). Based on this observation, one could expect putative SINEs from D. deserti to possess similar 5′-terminal sequences. Indeed, PCR with a primer specific to the 5′ end of B2 superfamily SINEs (B2, DIP, ID, and MEN) successfully amplified D. deserti genome regions containing SINEs. Interestingly, PCR was successful although the primer did not exactly match the 5′-terminal sequence of the identified SINE: three nucleotides in the central part of the primer differed from those in the D. deserti SINE consensus (Fig. 1). This similarity could suffice to initiate PCR; alternatively, priming could occur on SINE copies that matched the primer better.

It is worthy to mention an unexpected feature of single-primer PCR: SINEs were found at only one end in most PCR products, possibly due to occasional primer binding to non-SINE genomic sequences under non-stringent annealing conditions. This allows this approach to be used with moderately repetitive SINEs.

4.2. IDL-Geo, a complex SINE family from Geomyoidea

Using the method described above, a novel SINE family IDL-Geo was identified in the kangaroo rat genome. IDL-Geo proved to be specific only for two rodent families, Heteromyidea (pocket mice) and Geomyidea (pocket gophers); thus, most likely this SINE appeared before the divergence of these two families. IDL-Geo has a dimeric structure. Nine out of about 70 currently known SINE families are dimeric (Kramerov and Vassetzky, 2005) (Table 2). Five of them have one or both monomers derived from 7SL RNA. Three SINEs (CYN, Das-II, and Twin) consist of two similar tRNA-related monomers. Only IDL-Geo consists of two different monomers unrelated to 7SL RNA. The first tRNAAla(CGC)–derived monomer is similar to ID and other SINEs of the B2 superfamily (Fig. 2), while the origin of the second one is not known. The presence of box A of the pol III promoter suggests that it can also descend from a tRNA or another pol III-synthesized RNA. If so, the nucleotide sequence has undergone considerable changes and it is difficult to identify the evolutionary precursor of the Geo sequence.

While in both previously described ID-containing dimeric SINEs (B1-dID and B4), the other monomer is 7SL-derived, ID is combined with a different type element in IDL-Geo.

Generally, dimerization seems to render SINEs some yet unexplained advantages that enable them to more successfully propagate in the genome.

Acknowledgments

We are grateful to Drs. Ron DeBry (University of Cincinnati, USA), Francois Catzeflis (University of Montpellier II, France), Terry Robinson (University of Stellenbosch, South Africa) and Elena Lyapunova (Institute of Developmental Biology, Moscow, Russia) for providing DNA, tissues or animals, as well as to Drs. Sergey Morzunov (University of Nevada–Reno, USA) and Nikita Vassetzky for helpful discussions. This study was supported by the Russian Foundation for Basic Research (grant 05-04-49553).

References


