Mammalian Genome 13, 50–57 (2002).
DOI: 10.1007/s00335-001-2111-1

Mammalian Genome
Incorporating Mouse Genome

CAN—a pan-carnivore SINE family

Nikita S. Vassetzky, Dmitri A. Kramerov

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

Received: 13 July 2001 / Accepted: 18 September 2001

Abstract. Short retroposons or short interspersed elements (SINEs) constituting 5–10% genome have been isolated from various organisms. CAN SINEs initially found in American mink were named after dogs (Canis), and the range of their distribution in the genomes of carnivores and mammals in general remained topical. Here we demonstrate CAN sequences in representatives of all carnivore families, but not beyond carnivores, on the basis of sequence bank search and genomic PCR. Analysis of their distribution supports division of carnivores into caniform (dogs, mustelids, raccoons, bears, and pinnipeds) and feliform (cats, civets, and hyenas) lineages. CAN structure is considered in the context of their function and evolution.

A significant portion of the eukaryotic genome is composed of mobile elements propagated by retroposition, a process involving transcription and reverse transcription (Rogers 1985). Long retroposons (LINEs) code for the activities required for retroposition (reverse transcriptase and endonuclease), while short retroposons (SINEs) lack these.

SINEs are genomic repeats 80–400 bp long, apparently originating from RNA (more commonly tRNA); a typical SINE consists of three regions; a tRNA-related region, a tRNA-unrelated region, and an A-rich region. The tRNA-related region contains an internal promoter of RNA polymerase III and provides for its transcription (Daniels and Deininger 1985; Jagadeeswaran et al. 1981). Owing to the mechanism of amplification, the 3′-part of SINEs is A-rich, and the genomic elements are flanked by short direct repeats.

SINEs can be classified as members of several superfamilies sharing structural similarity, apparently inherited from a common ancestor [e.g., members of *Alu*/B1 superfamily originate from an ancient element FAM (Quentin 1992)]. The vast majority of SINE copies in the genome are incapable of active amplification, and the process is due to a few master sequences as indicated by the presence of distinct subfamilies and the activity of one or a few at any given time.

The process of SINE amplification in the genome goes within a certain period of evolution; for instance, SINEs of CORE superfamily amplified in the earliest days of vertebrate evolution and are inactive now (Gilbert and Labuda 2000), while *Alu* amplification is still going (Leeflang et al. 1992). No mechanism of specific SINE elimination from the genome is known, and SINEs remain in the genomes of successors: MIR is found in all mammals, while *Alu* is specific for primates, which allows their use as independent phylogenetic markers (Kramerov et al. 1999; Murata et al. 1993; Shimamura et al. 1997).

SINEs were initially found in primates and rodents (Krayev et al. 1980; Rubin et al. 1980); now it is generally accepted that several SINE families are present in most (probably all) higher eukaryotes. Usually, one SINE family has over 100,000 copies per haploid genome, while others are less abundant [e.g., the mouse genome contains ~100,000 B1s and B2s each (Kramerov et al. 1979) and ~40,000 IDs (Kass et al. 1996)].

In carnivores CAN short retroposons were first found in American mink (Lavrent'eva et al. 1989), then in the dog (Minnick et al. 1992) and harbor seal (Coltman and Wright 1994). This was a typical tRNA-related SINE family highly repeated in the genomes [~400,000 per dog haploid genome (Bentolita et al. 1999)]. For over a decade the taxonomic range of this SINE distribution remained controversial: most groups limited it to Canoidea (caniform carnivores including dogs, bears, raccoons, mustelids, skunks, seals, and walruses, while feliform carnivores include cats, civets, and hyenas) (Bentolila et al. 1999; Coltman and Wright 1994; Das et al. 1998; Zehr et al. 2001), whereas van der Vlugt and Lenstra (1995) and Precon Slattery et al. (2000) detected this SINE in the cat.

Here we resolve this dispute and demonstrate the presence of CAN SINEs in the genomes of all carnivore lineages, but not beyond carnivores. We also analyze the fine structure and arrange variants of CANs. Since this SINE family is now known in an unusually wide range of species, we consider phylogenetic relations of this short retroposon as well as the evolution of carnivores.

Materials and methods

We studied DNA of the following species. Carnivores: cat *Felis catus,* dog *Canis familiaris,* walrus *Odobenus rosmarus,* civet *Viverra zibetha;* non-carnivores: human *Homo sapiens,* mouse *Mus musculus,* Daurian hedgehog *Mesechinus* (*Hemiechinus*) *dauuricus,* water bat *Myotis daubentoni,* rabbit *Oryctolagus cuniculus,* short-eared elephant shrew *Macroscelides proboscideus,* cow *Bos taurus,* nine-banded armadillo *Dasypus novemcinctus,* Bennett's tree kangaroo *Dendrolagus bennettianus,* tree shrew *Tupaia glis;* and birds: gull *Larus cachinans,* white-naped crane *Grus vipio,* and chicken *Gallus gallus.*

The walrus, kangaroo, and elephant shrew tissues were provided by S. Popov (Moscow Zoo); the civet and tree shrew tissues were provided by O. Likhnova (Institute of Ecology and Evolution, Moscow), and the armadillo tissue was provided by R. DeBry (University of Cincinnati). For other DNA sources and isolation technique, see Borodulina and Kramerov (1999).

PCR conditions were described elsewhere (Kramerov et al. 1999; Serdobova and Kramerov 1998). The amplification products were analyzed by electrophoresis in 4% NuSieve (FMC) agarose gel. We used the following primers corresponding to the polymerase III promoter box A and the extension of tRNA-related region in CAN SINE: 5′-CTGGGTGGCTCAGTCRGT-3′ and 5′-AGCACAGAGCCYGAYGYG-3′. Under these PCR conditions, 100 ng genomic DNA sufficed to reveal single genomic copies;

Correspondence to: N.S. Vassetzky; E-mail: nvas@genome.eimb.ru

Table 1.

SUPER-ORDER	FAMILY	Species		No. of known CANs
Feliformia	Felidae (cats)	*Felis catus*—domestic cat		19 (+30)
	Hyaenidae (hyenas)	*Crocuta crocuta*—spotted hyena		0 (+1)
	Viverridae (civets)	*Cryptoprocta ferox*—fossa		0 (+4)
		Viverra zibetha—large Indian civet		PCR⊕
Caniformia	Candidae (dogs)	*Canis familiaris*—dog		79 (+9)
		Canis lupus—wolf		1 (=0)
		Vulpes vulpes—red fox		2 (+0)
	Mustelidae (mustelids)	*Mustela vision*—mink		8 (+7)
		Mustela putorius—polecat		1 (+0)
		Mustela frenata—long-tailed weasel		1 (0)
		Lontra longicaudis—river otter		1 (+0)
		Martes americana—marten		0 (+5)
		Lutra lutra—common otter		1 (+1)
		Taxidea taxus—American badger		0 (+2)
		Mephitis mephitis—striped skunk		1 (+0)
		Spilogale putorius—spotted skunk		1 (+0)
	Procyonidae (raccoons)	*Procyon lotor*—raccoon		1 (+0)
		Potos flavus—kinkajou		1 (+0)
	Ursidae (bears)	*Ailuropoda melanoleuca*—giant panda		6 (+0)
		Tremarctos ornatus—spectacled bear		1 (+0)
		Ursus arctos—brown bear		1 (+0)
		Ailurus fulgens—Red Panda		1 (+0)
	Phocidae (true seals)	*Phoca vitulina*—harbor seal		2 (+2)
		Hydrurga leptonyx—leopard seal		1 (+2)
		Erignathus barbatus—bearded seal		1 (+0)
		Leptonychotes weddelli—Weddell's seal		0 (+2)
		Lobodon carcinophagus—crab-eater seal		0 (+1)
	Odobenidae (walruses)	*Odobenus rosmarus*—walrus		2 (+0)
	Otariidae (eared seals)	*Callorhinus ursinus*—Northern fur seal		1 (+0)
		Zalophus californianus—Californian sea lion		1 (+0)

accordingly, 0.1 and 100 ng of carnivore and non-carnivore DNA were used, respectively.

Sequence similarity search in nucleotide banks was carried out using FASTA algorithm (Pearson and Lipman 1988). Initially we used a consensus of several published Canoidea CAN sequences as a search pattern; later it was refined by the found sequences. Found full-length sequences at least 60% identical to the search pattern or at least 70% identical partial sequences were considered for inclusion in the alignments. Since many genomic sequences have polypyrimidine regions, A-rich regions, or RNA polymerase III promoter, the similarity with the CAN sequence beyond these regions was decisive.

The alignments were generated manually with GeneDoc software (Nicholas and Nicholas 1997). Consensus sequences were also produced by GeneDoc from alignments of both full-length and partial elements.

For phylogenetic analysis the flanking sequences, polypyrimidine region, and A-rich tail were excluded from the alignment. Cladograms were constructed by the maximum parsimony, maximum likelihood, and neighbor joining methods by using the bootstrap procedure (100 replications) with a randomized order of sequences (100 times) by PHYLIP (version 3.57c) (Felsenstein 1993) and fastDNAmL (Olsen et al. 1994) software.

The number of SINE copies in the genome was roughly estimated from the proportion of CAN sequences in bank sequences for a given taxon using equation $N = n \times 3 \times 10^9/L$, where n is number of found CANs in the bank sequences of total length L; 3×10^9 states for the length of haploid mammalian genome.

The following CAN sequence designations were used. Cats (Felidae): Fel-01 and Fel-02, AB018479; Fel-03, AF039137; Fel-12, AF130593; Fel-21, AF130655; Fel-51, AF130700; Fel-52, AW646738; Fel-53, AW646884. Hyenas (Hyaenidae): Hya-01, AF180491. Bears (Ursidae): Urs-01, AF039738; Urs-02, AF039740; Urs-03, AF039741; Urs-04, AF076486; Urs-05, AF076488; Urs-06, AF076489; Urs-07, AF076492; Urs-08, AF076493. Raccoons (Procyonidae): Pro-01, AF039736; Pro-02, AF039737; Pro-03, AF039739. Mustelids (Mustelidae): Mus-01, AF039734; Mus-02, AF039735; Mus-03, AF072130; Mus-04, AF132101; Mus-05 and Mus-06, AF163863-1; Mus-08, MVB2RPT; Mus-09, MVISAA1A; Mus-10, MVITI1; Mus-11, MVU87245; Mus-13, AF306948; Mus-14, AF306949. Eared seals (Otariidae): Ota-01, AF039744; Ota-02, AF039745. Walruses (Odobenidae): Odo-01, AF039743. True seals (Phocidae): Pho-01, AF039742; Pho-04, PVSINEC; Pho-05, PVSINEE. The complete list of CAN SINEs as well as their alignments and consensus sequences are available on request.

We used published consensus sequences of rodent B2 (Bains and Temple-Smith 1989) and rat ID (Kim et al. 1994), while those for MEN and DIP were generated from the available sequences. The tRNAs similar to CAN and B2-related SINEs include human lysine $tRNA^{CTT}$ on Chr 5 (205394467–205394395) (identical to 10 more tRNA sequences); alanine $tRNA^{ACG}$, #63 on Chr 6 (32793728–32793799); and cysteine $tRNA^{GCA}$, #20 on Chr 17 (41257029–41256957) (Lowe 2001).

Results

Search for CAN-related sequences in the genomes of vertebrates. Sequence bank search for CAN-related sequences reveals full-length sequences at least 60% (typically more) identical to the consensus compiled from published CAN sequences in all carnivore lineages. Thus, we have found the following: dogs, 82 full-size and 9 incomplete elements; cats, 19 full-size and 30 incomplete elements; mustelids, 14 full-size and 13 incomplete elements; bears, 8 full-size elements; true seals, 5 full-size and 6 incomplete elements; raccoons, 3 full-size elements; civets, 4 incomplete elements; walruses and eared seals, 2 complete elements each; and hyenas, 1 nearly complete element (Table 1). Despite the high number of available canine sequences, they can be clearly divided into structural variants A, B, and C (data not shown). Alignment of the corresponding consensus sequences (Can-A—Can-C) complemented by full-size and not very unusual individual CAN sequences from other carnivore families is presented in Fig. 1.

The search for CAN in the available sequence banks of non-carnivore mammals, reptiles, amphibians, and fish yielded no such SINEs. Most found sequences contained polypyrimidine tract; the most similar SINEs found among non-carnivore genomic sequences belong to the B2 superfamily.

Surprisingly, 2 full-size and 15 incomplete elements with typical CAN structure were found among chicken genomic sequences (GGZ95203, GG5CHOXA, GGZ95160, GGZ95164, GGZ95165, GGZ95168, GGZ95170, GGZ95179, GGZ95181, GGZ95186, GGZ95187, GGZ95191, GGZ95192, GGZ95196, GGZ95197, GGZ95201, and GGZ95202; all but one were annotated as "microsatellites"). These avian sequences were markedly more similar to the feliform CAN than to caniform (and this similarity was higher than that between caniform and feliform CAN; chicken and feliform elements formed a clade in phylogenetic trees plotted by all tested methods (data not shown).

In order to test this striking finding, we carried out PCR with a variety of mammalian and avian genomic DNA with the primers selected not to reveal tRNA genes. The CAN-related sequences were revealed in all tested carnivore genomes (Fig. 2; lanes 2–5), but neither in non-carnivore mammals (the main lineages were selected; lanes 6–15) nor in birds (including chicken; lanes 16–18). Note that the amount of non-carnivore DNA (1000 times that of carnivore DNA) sufficed for detection of a single genomic copy.

PCR revealed the 5′-region of CAN in the civet genome (Fig. 2; lane 5); this completes the sequence data and confirms the presence of a typical CAN in civets.

Structure of CAN short retroposon. CAN is a typical tRNA-related short retroposon. It is usually about 170 bp long; however, the length ranges from below 150 to above 280 owing variability of the polypyrimidine region and A-rich tail, as well as extended insertions/deletions. Similar to other retroposons, CAN elements are flanked by short direct repeats of host DNA indicating a retroposition event (underlined in Fig. 1). The tRNA-related region contains putative boxes of RNA polymerase III promoter (Fig. 1; boxes A and B); however, there is an 11-nucleotide insertion between the boxes relative to the tRNA sequence (Fig. 1; ins 2).

The 3′-part of CAN contains a polypyrimidine region (a $(TC)_n$ motif) and an A-rich tail with polyadenylation signal AATAAA and RNA polymerase III terminator TTTT or TCTTT found in many CAN copies.

As with other SINEs, individual CAN sequences vary considerably. On the one hand, this is due to multiple single-nucleotide substitutions and indels. Note that a significant portion of such variability results from methylation-mediated CG sites mutation to TG or CA (Bird 1980).

On the other hand, one can note several larger rearrangements. Some of these apparently result from duplication of short sequences; their remains can be recognized as degenerate direct repeats (designated as DR1, DR2, and DR3 in Fig. 1). DR1 ($CTCAG_{2-3}TCNG$) is present in virtually all CAN sequences [the only exception out of all CAN sequences has other rearrangements in the considered region (Fig. 1, Fel-12), suggesting that it is secondary to the duplication]. Note that this repeat is less degenerate in the available feliform sequences. DR2 (TGTGCTG) following the tRNA-related region is present in about half of feline sequences; it is not found in the available sequences of other Feliformia. Actually, caniform CANs have slightly different sequences

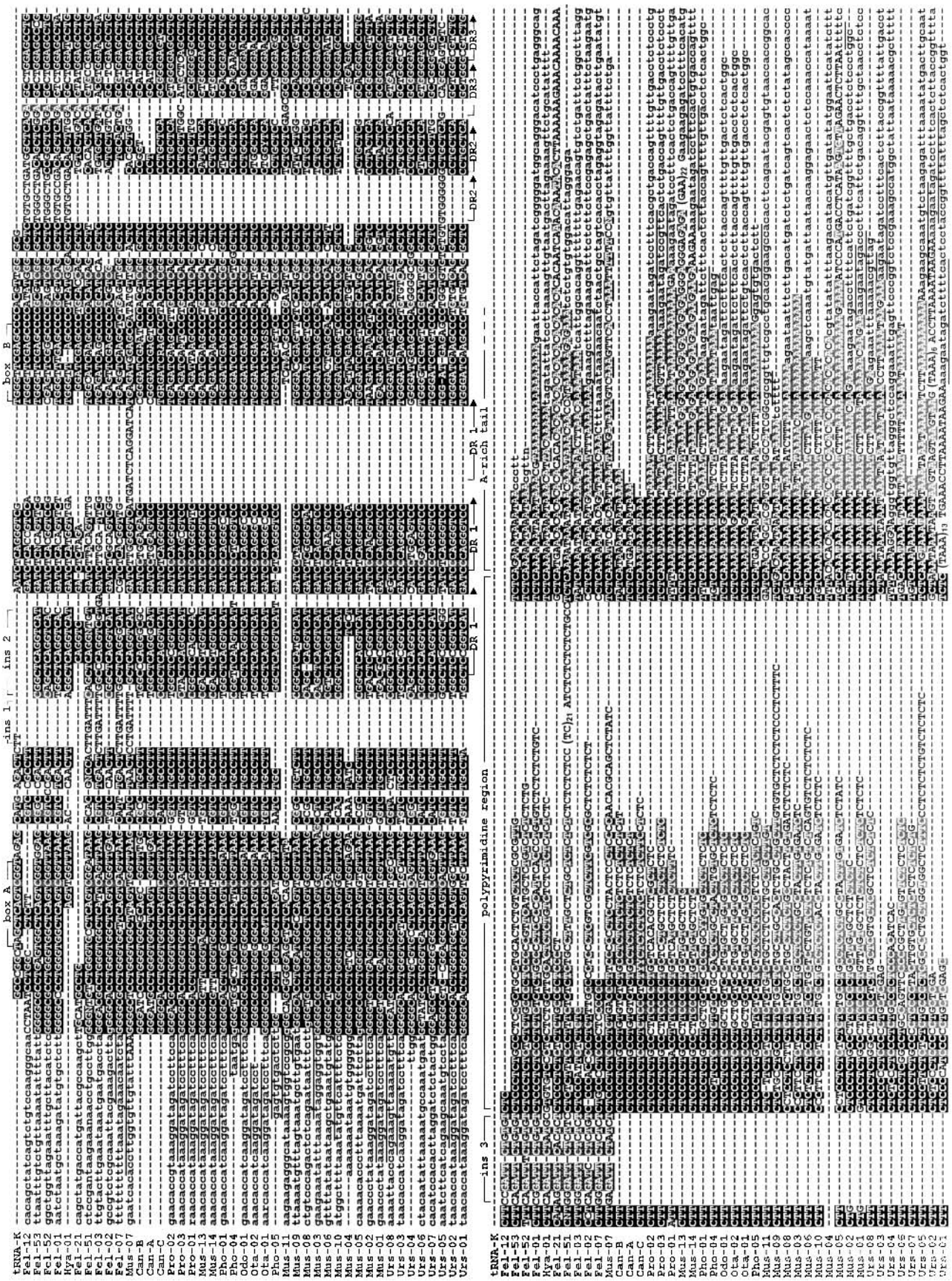

Fig. 1. Aligned nucleotide sequences of full-size CAN elements from carnivores. Only the structural variants (CAN-A—CAN-C) are given for Canidae. The retroposon sequences are given in upper case while the flanking host sequences are in lower case with the direct repeats underlined. Human lysine tRNA is included for reference. Internal duplications are shown with arrows below the sequences, while presumable RNA polymerase III promoters (A and B boxes), polypyrimidine region, A-rich tail, and insertions are indicated above. For other designations see text.

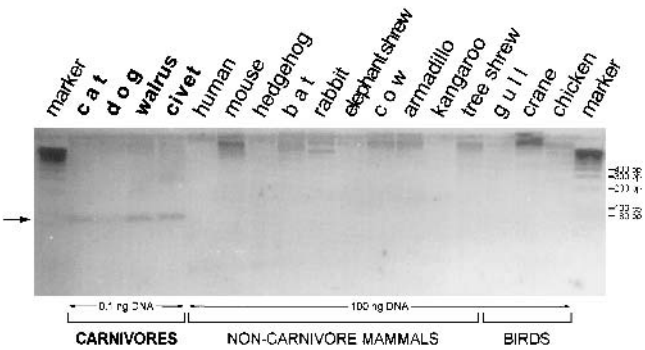

Fig. 2. CAN SINE PCR of genomic DNA from carnivore and non-carnivore vertebrates. Common species names are given above the lanes; the first and last lanes are size markers. The amount of genomic DNA in the reaction is indicated below; the arrow points to the CAN-specific band.

there (CCT, CCGGT, and CCTGCTC in the variants A–C, respectively). DR3 (GAGCCTG) can also be found in ~3/4 feline sequences (and in a half of civet CANs), while only remains of this repeat (or an independent insertion) can be found in Caniformia (—GCNGG). Another kind of rearrangement is presented by two insertions (ins 1 and ins 3 in Fig. 1). Ins 1 (TTGATTT) is found in ~1/3 feline, ~1/10 canine, and single otter (Mus-07) CANs, while ins 3 (TTGATTT) is specific for all available feliform CANs and a single caniform CAN (Mus-07).

The pattern of variability changes along the sequence; the tRNA-related region is more conserved and the RNA polymerase III promoter is most conserved in this region, while the polypyrimidine region and A-rich tail in the tRNA-unrelated part are extremely variable (Fig. 1).

Number of CAN copies in carnivore genomes. Our estimation of CAN copy number in the genomes based on their frequency in the sequence bank yields similar results for carnivore families: 300,000 in mustelids, 200,000 in cats and true seals, and 150,000 in dogs and bears. The samples for other families are too small. Although these estimations are very rough, our results correspond to a very accurate determination of CAN SINEs in dog genome (400,000) (Bentolila et al. 1999) as well as other estimations [seal, 10^5–10^6 (Coltman and Wright 1994); mink, 1–2 × 10^5 (Lavrent'eva et al. 1989)].

Phylogenetic relations between CAN SINEs. Sequence comparison with human tRNA database (Lowe 2001) demonstrates top similarity (70–79%) of all carnivore families CAN consensus sequences to lysine tRNAs (both TTT and CTT).

In order to evaluate phylogenetic relations between CAN sequences, we built cladograms, using three approaches (maximum parsimony, maximum likelihood, and neighbor joining) from the alignments of full-size individual sequences (excluding ones with extended deletions and rearrangements) and consensus sequences without the variable region (polypyrimidine region and A-rich tail). In some cases the alignments were truncated by the beginning of incomplete sequences in order to include hyena and civet sequences in the analysis. Figure 3 presents a cladogram built from the alignment of some complete individual CAN sequences and canine consensus sequences.

The topology of cladograms as well as bootstrap values of the nodes were not identical for different alignments and methods. However, several common features could be noted. In virtually all cladograms, the feliform CANs form a clade separate from the caniform sequences except for the otter Mus-07 (Fig. 3). Most pinniped (Phocidae, Odobenidae, and Otariidae) CANs also tend to cluster on the phylogenetic trees. Individual CANs of other carnivore families are intermixed.

Phylogenetic relations of CAN and other SINEs. Initially, CAN SINEs were found by similarity with rodent B2 (Lavrent'eva et al. 1989). We also tried to assess the relations between B2 superfamily and CAN. B2 superfamily includes rodent SINEs B2, ID, DIP, and MEN (Serdobova and Kramerov 1998). A sequence comparison of B2, ID, DIP, and MEN consensus sequences with human tRNA database (Lowe 2001) (excluding 'pseudo' sequences) demonstrates top similarity with alanine (ID, 80%; MEN, 81%) and cysteine (B2, 71%; DIP, 65%) rather than lysine tRNAs. The alignment of B2 superfamily and CAN consensus sequences together with alanine and lysine tRNAs (Fig. 4) demonstrates two clearly distinct groups of sequences (B2-like and CAN); similarity between these groups does not go beyond the similarity between the related tRNAs (Ala/Cys and Lys), which is most pronounced in the regions outside the A and B boxes. No significant similarity was found between B2-like and CAN sequence in the tRNA-unrelated region except for DIP element also carrying polypyrimidine region (Fig. 4).

Discussion

Distribution of CAN SINEs in animal genomes. Full-size CAN sequences were found in the genomic banks of representatives of all carnivore families except civets, where only incomplete sequences were found. The PCR experiments confirm the presence of CAN in the genomes of carnivores and civets in particular, since the 5′-region missing in the found civet sequences was amplified.

Sequence search yielded no CAN in a wide range of non-carnivore species except for chicken. However, our PCR data indicate the absence of CAN in non-carnivore genomes. We cannot explain the presence of CAN-related sequences in the chicken sequence bank. Since we failed to detect CAN in avian genomes by PCR, we should rather attribute these sequences to a cloning artifact. Indeed, no such mosaic distribution of short retroposons in host genomes has been documented, and the presence of CAN SINEs in carnivores and birds but no other organisms should be very surprising.

Taken together, our data indicate the presence of CAN short retroposons in the genomes of all carnivore families but not beyond this order. Apparently, previous failures to detect CAN in Feliformia were due to methodical limitations (most attempts were based on DNA hybridization), while databank search and PCR proved to be more sensitive and reliable.

Our estimates of CAN abundance in carnivore genomes (1.5–3 × 10^5 per haploid genome) as well as most published estimates (10^5–10^6) suggest that CAN is the major short retroposon of carnivores. This agrees with conclusions based on comprehensive analysis of dog genomic repeats (Bentolila et al. 1999).

Structure of CAN short retroposon. Although CAN has a typical structure of a tRNA-related short retroposon, there are some peculiarities, including an insertion ins 2 and polypyrimidine region. The ins 2 insertion alters the distance between the RNA polymerase III promoter boxes from 29 bp in the tRNA to ~40 in the short retroposon (Fig. 1). Interestingly, the 5′-part of the insertion is quite similar to the A box; however, the last nucleotides most conserved in box A of short retroposons are different in ins 2.

The polypyrimidine region mostly composed of $(TC)_n$ motif is found in the tRNA-unrelated region of CAN. It is also specific for several other mammalian SINEs including rabbit C repeat (Cheng et al. 1984), rodent DIP (Serdobova and Kramerov 1998), bat VES (Borodulina and Kramerov 1999), and insectivore TAL, ERI-1,

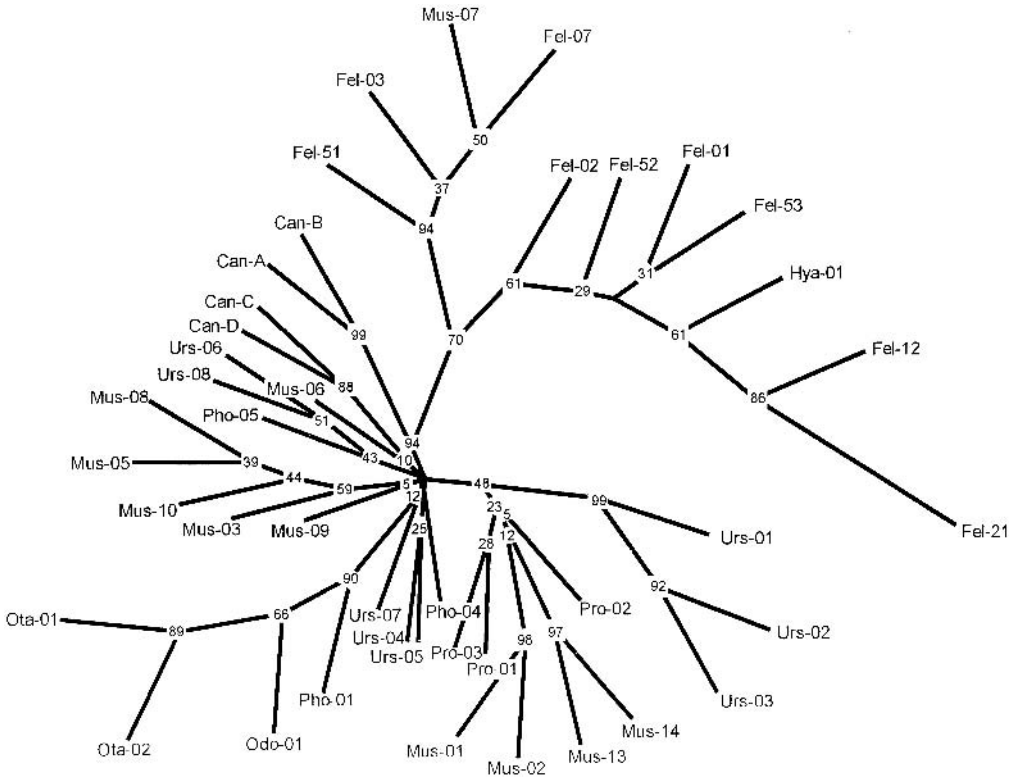

Fig. 3. Cladogram of CAN individual sequences and canine consensus sequences plotted by maximum parsimony method. Numbers indicate bootstrap values.

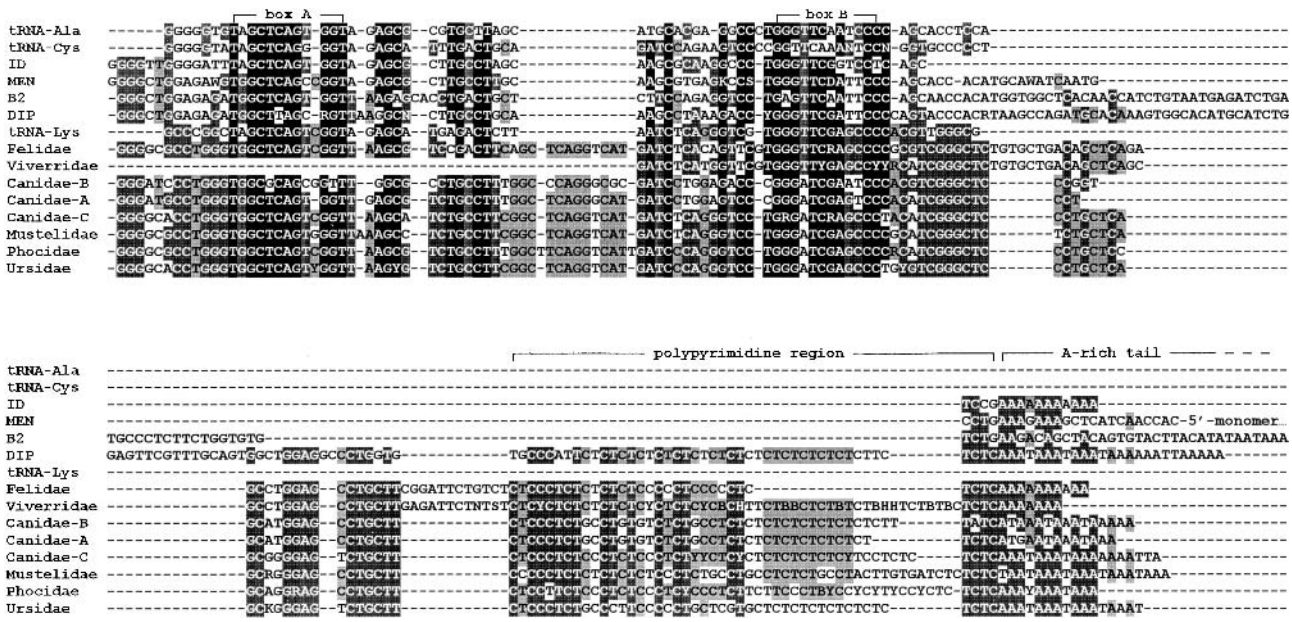

Fig. 4. Aligned nucleotide consensus sequences of B2-like and CAN SINEs as well as related tRNAs. For designations see Fig. 1.

and ERI-2 elements (Borodulina and Kramerov 2001). The functional significance of this structure remains unknown; however, its presence in unrelated mammalian SINEs is amazing.

The polypyrimidine region is quite variable both in sequence and length (Fig. 1), possibly resulting from microsatellite amplification, which could also increase the length of this region in the ancestor master sequence. Note that the length of the polypyrimidine region varies in the CAN copies integrated in the same site (Pro-01, -02, -03; Mus-01, -02; Pho-01; Ota-01, -02; Odo-01; Urs-01, -02, -03, as well as wolf and red fox sequences AF039732 and AF039733 not included in Fig. 1) even in relatively close species (e.g., Mus-01 and -02), while the sequence of this region corresponds to the species relations (e.g., the 3′-part of the polypyrimidine region in raccoons and skunks can be distinguished from that of other carnivores). A similar event can be observed in the A-rich tail of skunk Mus-13 element; apparently the GAAGAA sequence in the 3′-terminal direct repeat became the target for amplification of $(GAA)_n$ microsatellite (Fig. 1); note the absence of this micro-

satellite in the same integration site in other organisms except for another skunk (Mus-14) where $(GAA)_5$ can be found. These observations suggest that such length variability in the tRNA-unrelated region can be owing to post-integration events, possibly related to microsatellite amplification as observed for primate Alu (Arcot et al. 1995).

In addition to the polypyrimidine region variability, CAN variants differ by numerous single-nucleotide substitutions as well as several large rearrangements (insertions and deletions), which is quite typical for short retroposons especially when a set of relatively distant host species is considered (e.g., Kramerov and Vassetzky 2001). Anyway, accurate analysis demonstrated higher variability of the polypyrimidine region as compared to the neighboring regions (both the SINE tRNA-related region and the host integration site) (Zehr et al. 2001).

The A-rich tail is quite typical and belongs to the T^+ class specific for most mammalian SINEs (Borodulina and Kramerov 2001). A tail of this class carries both terminator of transcription and AATAAA signals; by analogy with rodent B2 SINE (Kramerov et al. 1990), one can propose that CAN RNA polymerase III transcripts are polyadenylated.

Generally, the tRNA-unrelated region of short retroposons was proposed to originate from LINEs (Okada and Hamada 1997) that provide reverse transcriptase for SINE retroposition. However, we failed to find considerable similarity between CAN and known LINEs; this is also the case for most mammalian short retroposons and may confirm proposed relaxed specificity of RNA recognition for a mammalian LINE reverse transcriptase (Okada and Hamada 1997).

Evolution of CAN SINEs and carnivores. According to the popular master gene model, only a few highly active "master" short retroposons are capable of amplification, giving rise to non-propagating copies (Shen et al. 1991). Accordingly, several structural variants can be recognized for CANs. For instance, there are three clearly distinct variants of Canidae CANs. Generally, the distribution of these variants in the host genomes corresponds to the host evolution with a few exceptions; e.g., DR2 duplication is found in half of feliform CANs but not in caniform ones, but the insertion insI is found in the feliform CANs and a single caniform sequence Mus-07. A similar interspersion of variants was observed for other SINEs too; for instance, salmonid Hpa I family (Hamada et al. 1997; Takasaki et al. 1994). We believe that various ancestor variants were formed before their active amplification (and radiation of carnivores) so that later independent choice of a particular variant could correspond or not to the host evolution, and single copies of most variants can be found. Otherwise, horizontal transfer should be assumed, which has never been documented for SINEs and seems unlikely.

Recent analysis of a CAN at exactly the same integration site in transthyretin gene (Zehr et al. 2001) demonstrated CAN copies in all available caniform genomes but not in the feliform ones. This proves the generally accepted division of carnivores into Feliformia and Caniformia lineages, e.g., seals and dogs are related more to each other than either of them to cats. Our data on distribution of CAN among carnivores supports both the division into these lineages and monophily of carnivores.

In terms of short retroposons evolution from transfer and 7SL RNAs (or their pseudogenes), one can note a trend for elongation in their 3′-region realized as short tandem duplications [e.g., B1 (Labuda et al. 1991)], dimerization (Alu, galago SINE type II, MEN, B1-dlD (Kramerov and Vassetzky 2001 and references therein), or polypyrimidine regions. The deviation of SINEs from ancestor RNA in the 3′-region should be related either to the retroposition process (e.g., acquiring higher affinity to the reverse transcriptase) or to the control of transcription. Indeed, most of the time transcription of short retroposons is repressed except for special conditions such as early embryogenesis (Bachvarova 1988; Grigoryan et al. 1985) or stress (Liu et al. 1995; Panning and Smiley 1993).

Acknowledgments We thank Dr. Youri Kravatsky for his permanent help with computer hardware and software. This work was supported by the Russian Foundation for Basic Research (projects 99-04-49177 and 99-04-49178).

References

Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995) Alu repeats: a source for the genesis of primate microsatellites. Genomics 29, 136–44.

Bachvarova R (1988) Small B2 RNAs in mouse oocytes, embryos, and somatic tissues. Dev Biol 130, 513–523

Bains W, Temple-Smith K (1989) Similarity and divergence among rodent repetitive DNA sequences. J Mol Evol 28, 191–199

Bentolila S, Bach JM, Kessler JL, Bordelais I, Cruaud C et al. (1999) Analysis of major repetitive DNA sequences in the dog (*Canis familiaris*) genome. Mamm Genome 10, 699–705

Bird AP (1980) DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res 8, 1499–504

Borodulina OR, Kramerov DA (1999) Wide distribution of short interspersed elements among eukaryotic genomes. FEBS Lett 457, 409–413

Borodulina OR, Kramerov DA (2001) Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. Mamm Genome, in press

Cheng JF, Printz R, Callaghan T, Shuey D, Hardison RC (1984) The rabbit C family of short, interspersed repeats. Nucleotide sequence determination and transcriptional analysis. J Mol Biol 176, 1–20

Coltman DW, Wright JM (1994) Can SINEs: a family of tRNA-derived retroposons specific to the superfamily Canoidea. Nucleic Acids Res 22, 22,2726–2730

Daniels GR, Deininger PL (1985) Repeat sequence families derived from mammalian tRNA genes. Nature 317, 819–822

Das M, Chu LL, Ghahremani M, Abrams-Ogg T, Roy MS et al. (1998) Characterization of an abundant short interspersed nuclear element (SINE) present in *Canis familiaris.* Mamm Genome 9, 64–69

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package). University of Washington, Seattle

Gilbert N, Labuda D (2000) Evolutionary inventions and continuity of CORE-SINEs in mammals. J Mol Biol 298, 365–377

Grigoryan MS, Kramerov DA, Tulchinsky EM, Revasova ES, Lukanidin EM (1985) Activation of putative transposition intermediate formation in tumor cells. EMBO J 4, 2209–2215

Hamada M, Kido Y, Himberg M, Reist JD, Ying C et al. (1997) A newly isolated family of short interspersed repetitive elements (SINEs) in coregonid fishes (whitefish) with sequences that are almost identical to those of the Smal family of repeats: possible evidence for the horizontal transfer of SINEs. Genetics 146, 355–367

Jagadeeswaran P, Forget BG, Weissman SM (1981) Short interspersed repetitive DNA elements in eucaryotes: transposable DNA elements generated by reverse transcription of RNA pol III transcripts? Cell 26, 141–142

Kass DH, Kim J, Deininger PL (1996) Sporadic amplification of ID elements in rodents. J Mol Evol 42, 7–14

Kim J, Martignetti JA, Shen MR, Brosius J, Deininger P (1994) Rodent BC1 RNA gene as a master gene for ID element amplification. Proc Natl Acad Sci USA 91, 3607–3611

Kramerov DA, Vassetzky NS (2001) Structure and origin of a novel dimeric retroposon b1-did. J Mol Evol 52, 137–143

Kramerov DA, Grigoryan AA, Ryskov AP, Georgiev GP (1979) Long double-stranded sequences (dsRNA-B) of nuclear pre-mRNA consist of a few highly abundant classes of sequences: evidence from DNA cloning experiments. Nucleic Acids Res 6, 697–713

Kramerov DA, Tillib SV, Shumyatsky GP, Georgiev GP (1990) The most abundant nascent poly(A) + RNAs are transcribed by RNA polymerase ill in murine tumor cells. Nucleic Acids Res 18, 4499–4506

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

Krayev AS, Kramerov DA, Skryabin KG, Ryskov AP, Bayev AA, Georgiev GP (1980) The nucleotide sequence of the ubiquitous repeti-

tive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. Nucleic Acids Res 8, 1201–1215
Labuda D, Sinnett D, Richer C, Deragon JM, Striker G (1991) Evolution of mouse B1 repeats: 7SL RNA folding pattern conserved. J Mol Evol 32, 405–414
Lavrent'eva MV, Rivkin MI, Shilov AG, Kobets ML, Rogozin IB, et al. (1989) B2-like repetitive sequence in the genome of the American mink. Dokl Akad Nauk SSSR 307, 226–228
Leeflang EP, Liu WM, Hashimoto C, Choudary PV, Schmid CW (1992) Phylogenetic evidence for multiple Alu source genes. J Mol Evol 35, 7–16
Liu WM, Chu WM, Choudary PV, Schmid CW (1995) Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. Nucleic Acids Res 23, 1758–1765
Lowe T (2001) GtRDB: The Genomic tRNA Database http://RNA.wustl.edu/tRNAdb
Minnick MF, Stillwell LC, Heineman JM, Stiegler GL (1992) A highly repetitive DNA sequence possibly unique to canids. Gene 110, 235–238
Murata S, Takasaki N, Saitoh M, Okada N (1993) Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc Natl Acad Sci USA 90, 6995–6999
Nicholas KB, Nicholas HBJ (1997) GeneDoc: a tool for editing and annotating multiple sequence alignment. Distributed by the author
Okada N, Hamada M (1997) The 3′ ends of RNA-derived SINEs originated from the 3′ ends of LINEs: a new example from the bovine genome. J Mol Evol 44, 52–56
Olsen GJ, Matsuda H, Hagstrom R, Overbeek R (1994) fastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput Appl Biosci 10, 41–48
Panning B, Smiley JR (1993) Activation of RNA polymerase III transcription of human Alu repetitive elements by adenovirus type 5: requirement for the E1b 58-kilodalton protein and the products of E4 open reading frames 3 and 6. Mol Cell Biol 13, 3231–3244
Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85, 2444–2448
Precon Slattery J, Murphy WJ, O'Brien SJ (2000) Patterns of diversity among SINE elements isolated from three Y-chromosome genes in carnivores. Mol Biol Evol 17, 825–829
Quentin Y (1992) Origin of the Alu family: a family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. Nucleic Acids Res 20, 3397–3401
Rogers JH (1985) The origin and evolution of retroposons. Int Rev Cytol 93, 187–279
Rubin CM, Houck CM, Deininger PL, Friedmann T, Schmid CW (1980) Partial nucleotide sequence of the 300-nucleotide interspersed repeated human DNA sequences. Nature 284, 372–374
Serdobova IM, Kramerov DA (1998) Short retroposons of the B2 superfamily: evolution and application for the study of rodent phylogeny. J Mol Evol 46, 202–214
Shen MR, Batzer MA, Deininger PL (1991) Evolution of the master Alu gene(s). J Mol Evol 33, 311–320
Shimamura M, Yasue H, Ohshima K, Abe H, Kato H et al. (1997) Molecular evidence from retroposons that whales form a clade within eventoed ungulates. Nature 388, 666–670
Takasaki N, Murata S, Saitoh M, Kobayashi T, Park L et al. (1994) Species-specific amplification of tRNA-derived short interspersed repetitive elements (SINEs) by retroposition: a process of parasitization of entire genomes during the evolution of salmonids. Proc Natl Acad Sci USA 91, 10153–10157
van der Vlugt HH, Lenstra JA (1995) SINE elements of carnivores. Mamm Genome 6, 49–51
Zehr SM, Nedbal MA, Flynn JJ (2001) Tempo and mode of evolution in an orthologous Can SINE. Mamm Genome 12, 38–44