Mammalian Genome 9, 64–69 (1998).

Mammalian Genome
Incorporating Mouse Genome

Characterization of an abundant short interspersed nuclear element (SINE) present in *Canis familiaris*

Manjula Das,[1] Lee Lee Chu,[1] Majid Ghahremani,[1] Tony Abrams-Ogg,[2] Michael S. Roy,[3] David Housman,[4] Jerry Pelletier[1,5]

[1]Dept. of Biochemistry, McIntyre Medical Sciences Building, McGill University, Rm 902, 3655 Drummond, Montreal, Quebec, Canada, H3G 1Y6
[2]Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada, N1G 2W1
[3]Zoological Institute, University of Copenhagen, Dept. Population Biology, Univeristetsparken 15, DK-2100, Copenhagen, Denmark
[4]Center for Cancer Research, Massachusetts Institute of Technology, 40 Ames St., Cambridge, Massachusetts 02139, USA
[5]Dept. of Oncology, McGill University, 3655 Drummond St., Montreal, Quebec, Canada, H3G 1Y6

Received: 23 July 1997 / Accepted: 9 September 1997

Abstract. A short interspersed nuclear element (*Can* SINE) of ~130–150 bp was cloned and characterized from *Canis familiaris*. We demonstrate that this element is interspersed, present approximately every 5–8.3 kbp, and many are sufficiently close to allow IRS (interspersed repetitive DNA sequences) PCR. Sequence analysis of >20 *Can* SINEs from the dog has identified a conserved region that was used to design oligonucleotides for IRS PCR. Since *Can* SINEs are not present in human or rodent genomes, IRS PCR using oligonucleotides directed to the conserved region of *Can* SINEs can be used to simplify analysis of canid DNA in somatic cell hybrids, as well as in large insert cloning vectors. We demonstrate that the canid IRS products are polymorphic and could be developed as genetic markers for filter-based genotyping in this organism.

Introduction

The wealth of genetic conditions in the domestic dog, *Canis familiaris,* holds considerable promise for understanding basic and applied problems in mammalian biology. The dog is a particularly attractive model for studying congenital disorders (for example, see http://probe.nalusda.gov:8300/cgi-bin/browse/omia), behavioral diversity, sporadic malignancies, and pharmacogenetics, since many of these phenotypes or traits are not accurately represented in current popular small-animal models. Accessing this genetic information in the dog will require the development of a detailed physical and genetic map of the genome. The development of canine simple sequence length polymorphisms (SSLPs) represents a first step in such a task (Ostrander et al. 1993, 1995; Francisco et al. 1996; Yuzbasiyan-Gurkan et al. 1997). Recently, whole genome scans employing 213 SSLPs have identified a marker linked to the canine copper toxicosis locus in Bedlington Terriers (Yuzbasiyan-Gurkan et al. 1977). Currently, these markers are of limited usefulness since they are unmapped, although the application of fluorescent in situ hybridization (FISH) techniques to canine genome studies promises to quickly remediate this situation (Fischer et al. 1996).

The development of IRS (interspersed repetitive DNA sequences) PCR (polymerase chain reaction) has alleviated many difficulties associated with genetic mapping in human and mouse genomics, making possible the specific amplification of DNA from complex sources without prior sequence knowledge. This has greatly simplified analysis of human DNA within somatic cell hybrids, yeast artificial chromosomes (YACs), cosmids, and phages (Nelson et al. 1989; Ledbetter et al. 1990). This technique is based on the observation that unique genomic sequences between two repeat elements can be amplified by using repeat element-specific primers, and has been applied to the isolation of chromosome region-specific polymorphic markers (Guzzetta et al. 1991; Brooks-Wilson et al. 1992), identification of chromosome-specific YACs (Chumakov et al. 1992), contig assembly (Zucman et al. 1992; Hunter et al. 1994; Liu et al. 1995), linkage mapping (Zietkiewicz et al. 1992), and high-resolution, filter-based genotyping (McCarthy et al. 1995). In this report, we have extended IRS PCR to canine genomics. We have characterized a short interspersed nuclear element (SINE) from *Canis familiaris* and demonstrate the feasibility of IRS PCR in this species.

Correspondence to: J. Pelletier

Materials and methods

Materials and general methods. Restriction endonucleases, calf intestinal alkaline phosphatase, the Klenow fragment of DNA polymerase I, and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). [^{35}S]-deoxyadenosine 5′-(α-thio)thiphosphate (1000–1500 Ci/mmol), [α-^{32}P]-deoxyctidine 5′-triphosphate (3000 Ci/mmol), and [γ-^{32}P]-adenosine triphosphate (3000 Ci/mmol) were obtained from New England Nuclear.

Preparation of high-molecular-weight genomic DNA, plasmid DNA, restriction enzyme digestions, agarose gel electrophoresis of DNA, DNA ligation, and bacterial transformations were carried out by standard methods (Sambrook et al. 1989, and references therein). Subclones of PCR amplifications were sequenced by the chain termination method (Sanger et al. 1977) with double-stranded DNA templates.

Generation and screening of small insert canine genomic library. MDCK cells, derived from an apparently normal adult female cocker spaniel (American Type Culture Collection), were maintained in Eagle's MEM with Earle's BSS supplemented with 10% heat-inactivated fetal calf serum, glutamine, penicillin, and streptomycin. For the generation of small insert canine genomic libraries, high-molecular-weight canine genomic DNA was prepared from MDCK cells and digested with *Sau*3AI, *Rsa*I, or *Alu*I. Restriction enzyme-digested genomic DNA was fractionated on a 1.5% agarose gel, and fragments <300 bp were purified by the glass milk method (Qiagen, Mississavga, Ontario). Gel-purified DNA was then ligated to pBluescript KS II+ (Stratagene, Aurora, Ontario) which had been linearized with either *Bam*HI (for *Sau*3AI-digested DNA) or *Sma*I (for *Alu*I-and *Rsa*I-digested DNA). Following transformation into *E. coli* DH5α, clones were picked, replicated, and prepared for colony hybridization by a modified Benton and Davis (1977) approach. Filters were prehybridized in Church's buffer (0.5 M NaH_2PO_4 [pH 7.0], 1 mM EDTA, 7% SDS, 0.5% BSA) at 68°C for 30 min, then hybridized at 68°C for 12 h with total canine genomic DNA that had been radiolabeled with [α-^{32}P]-dCTP by the random primer method (10^6 cpm/ml; Sambrook et al. 1989). Filters were washed once in 2 × SSC (20 × SSC is 3.0 M NaCl/0.3 M trisodium citrate)/0.5% SDS at 68°C for 20 min, once in 0.2 × SSC/0.1% SDS at

68°C for 20 min, and once in 0.1 × SSC/0.1% SDS at 68°C for 20 min. After exposure to X-Omat film (Kodak) at −70°C for 12 h with an intensifying screen, positive clones were scored and picked from the master plates for sequence analysis.

Canine phage genomic library screening. A canine phage genomic library (Clontech Palo Alto, CA) was plated at low density (~100 plaques/15 cm^2 dish). Following transfer of plaques to nitrocellulose filters, filters were processed as previously described (Sambrook et al. 1989). Prehybridization(6 × SSC/5 × Denhardt's/0.05 sodium pyrophosphate/100 μg/ml sonicated salmon sperm DNA/0.5% SDS) was performed at 37°C for 1 h. Hybridization (6 × SSC/1 × Denhardt's/100 μg/ml yeast tRNA/0.05% sodium pyrophosphate) was performed with oligonucleotide K9-AS at 37°C overnight with a probe concentration of 10^6 cpm/ml. Filters were washed in6 × SSC/0.05% sodium pyrophosphate three times at room temperature for a total of 15 min and once at 45°C for 15 min.

DNA was purified from ten randomly selected genomic phage clones by the plate lysate method (Sambrook et al. 1989). Following restriction enzyme digestion and fractionation on an 0.8% agarose gel, DNA was transferred to Hybond-N+ according to the manufacturer's recommendations (Amersham). The filter was treated for hybridization as described above and probed with radiolabeled K9-S (10^6 cpm/ml) at 37°C for 1 h. Washes were performed in 2 × SSC/0.05% SDS at room temperature for 30 min, with several changes, twice in 0.1 × SSC/0.1% SDS at room temperature for 40 min, and once in 6 × SSC/0.05% sodium pyrophosphate at 48°C for 30 min.

IRS PCR. Conditions for IRS PCR were essentially as previously described (Nelson et al. 1989). Essentially, PCRs were carried out in a total volume of 100 μl of 1 μg of DNA, primer at 1 μM in 50 mM KCl/10 mM Tris-HCl [pH 8.0]/1.5 mM $MgCl_2$/0.01% gelatin/300 μM dATP/300 μM dCTP/300 μM dGTP/300 μM dTTP (Pharmacia), and 2.5 units of Thermus aquaticus polymerase (Gibco-BRL) for 35 cycles of 94°C denaturation (1 min), 55°C annealing (45 s), and 68°C extension (5 min) in an automated thermal cycler (MJ Research).

For cloning of IRS PCR products, PCRs were fractionated on a 1.5% agarose gel, and products around the 200-bp molecular mass range were excised and purified by glass beads (Qiagen). This size-fractionated material was then ligated to pKS II+-based T-modified vector. Following transformation into DH10β and X-gal selection for recombinant plasmids, the sizes of the inserts were estimated from mini-preps. Clones were sequenced with T3 and universal primers to determine the sequence of the IRS product.

For polymorphism detection, IRS products from a number of breeds (see Fig. 5B) were fractionated on an 8% non-denaturing SSCP gel (50:1 acrylamide/bisacrylamide) with the electrophoresis being carried out at 4°C at constant power (30 Watts) for 4.5 h. Samples were electroblotted to Hybond N+ in 0.6 × TBE buffer (1 × TBE is 0.09 M Trisborate [pH 8.0], 0.002 M EDTA) at 40 volts for 2 h at 4°C. The blot was prehybridized in ExpressHyb (Clontech) for 30 min and hybridized with oligonucleotides (10^7 cpm/ml) in ExpressHyb at 37°C for 1 h. Washes were performed with 2 × SSC/0.05% SDS at RT for 1 h with several changes and with 0.2 × SSC/0.1% SDS at RT for 1 h with one change, after which the blot was exposed to X-OMAT (Kodak) film at −70°C overnight with two intensifying screens.

Results

Genetic and physical mapping in the dog with IRS PCR will depend on the use of PCR primers to amplify DNA sequences located between two interspersed repetitive elements oriented in opposite directions in the genome and separated by a distance that is within range of a conventional PCR. Individual loci can then be assayed by sampling the single-copy sequence located between the pairs of repetitive elements with a hybridization approach to distinguish among various polymorphic alleles. In humans and mice, this method takes advantage of conserved sequence elements within *Alu* (Deininger et al. 1981; Jurka and Smith 1988; Nelson et al. 1989) or B1/B2 (Hastie 1989) repeats, respectively, to design PCR primers that can extend from a large number of target sites. To eventually apply the same approach to canine genomics, we sought to better characterize *Can* SINEs in the dog genome and identify areas of sequence conservation to implement IRS PCR.

Identification and characterization of a canid repetitive element. Three small insert genomic libraries from *C. familiaris* were screened with radiolabeled total dog genomic DNA to identify clones containing high copy repetitive DNA elements. Following sequencing, related elements from genomic clones were visually identified by the presence of $(CT)_n$ and poly $(A)_n$ tracts at one end (arbitrarily defined as the downstream boundary; Figs. 1 and 2). Owing to the presence of 3 *Sau*3AI recognition sequences within many *Can* SINE members, clones from the *Sau*3AI library yielded sequence information for only the downstream half of the SINE (Fig. 1; clones are indicated by a # sign). Clones from the *Alu*I and *Rsa*I libraries yielded sufficient sequence information to establish upstream and downstream boundaries of the SINE. Sequence alignment revealed that the majority of genomic clones analyzed (>95%) contained a single repetitive element of the same family (Fig. 1). The *Can* SINE element identified is ~130–150 bp (taking into account the CT repeat and poly A tail tract) and from the same family as three previously identified sequences found in the dog genome and whose consensus sequence (REPDNA) is presented in Fig. 1 for comparison (Minnick et al. 1992). A BLAST search of the nonredundant nucleotide database at NCBI (National Center for Biotechnology Information) with clone R23 as the query revealed at least eight dog genes harboring similar SINEs, not all of them complete in structure, however (Fig. 1). A consensus SINE derived from alignment of these sequences is presented in Fig. 1. In addition, related sequences are also present in the mink (Lavrentiea et al. 1989) and harbor seal genome (Coltman and Wright 1994; Fig. 1, MVB2RPT and PVSINEC).

As previously noted (Minnick et al. 1992; Coltman and Wright 1994), two potential internal polymerase III control regions are present within the consensus SINE—a putative A box, similar to the polymerse IIA box consensus sequence (5′TGGCNNAGTNGG3′) and a B box, similar to the 5′GGTTCGANNCC3′ box B consensus sequence (Deininger 1989). Comparative sequence analysis among the genomic clones revealed a higher degree of conservation within the downstream half of the *Can* SINE compared with the upstream half (Fig. 1). To develop a probe that would allow us to target many members of this SINE family in PCRs, we designed oligonucleotides to a segment within the conserved downstream region (Fig. 1; K9-A and K9-AS).

Short-period interspersed elements in eukaryotes are generally flanked by direct repeats (reviewed in Jelinek 1982). In addition, an A-rich region is characteristically present immediately downstream of the SINE. To determine if any of the canid repeats we characterized harbored direct repeats, genomic sequence flanking several of our clones (as well as four genes from the NCBI database) was determined and inspected (Fig. 2). All elements, with the exception of the one identified in the p53 gene, contain $(CT)_n$ tracts and an A-rich region downstream of the repeat element (Fig. 2). In fact all of the genomic clones we have isolated and presented in Fig. 1 contain this identifier feature downstream of the repeat (J. Pelletier; data not shown). All SINE clones, in which extensive downstream sequence was obtained, were found to be flanked by direct repeats varying in length from 8 to 17 bp. These elements were not conserved among family members, but were unique to each canid SINE. These features are consistent with a model by which the direct repeats flanking each SINE have resulted from the duplication of a unique DNA sequence at the target site of insertion into chromosomal DNA as proposed for other eukaryotic SINEs.

Repetition frequency and dispersed nature of the Can SINE. To establish an estimate of the spacing of short-period repeats in the canine genome, ten phage clones were picked at random from a λ FIXII canine genomic library (average size ~15 kbp). DNA was isolated from the clones and digested with *Bam*HI. Following fractionation by agarose gel electrophoresis, digests were transferred

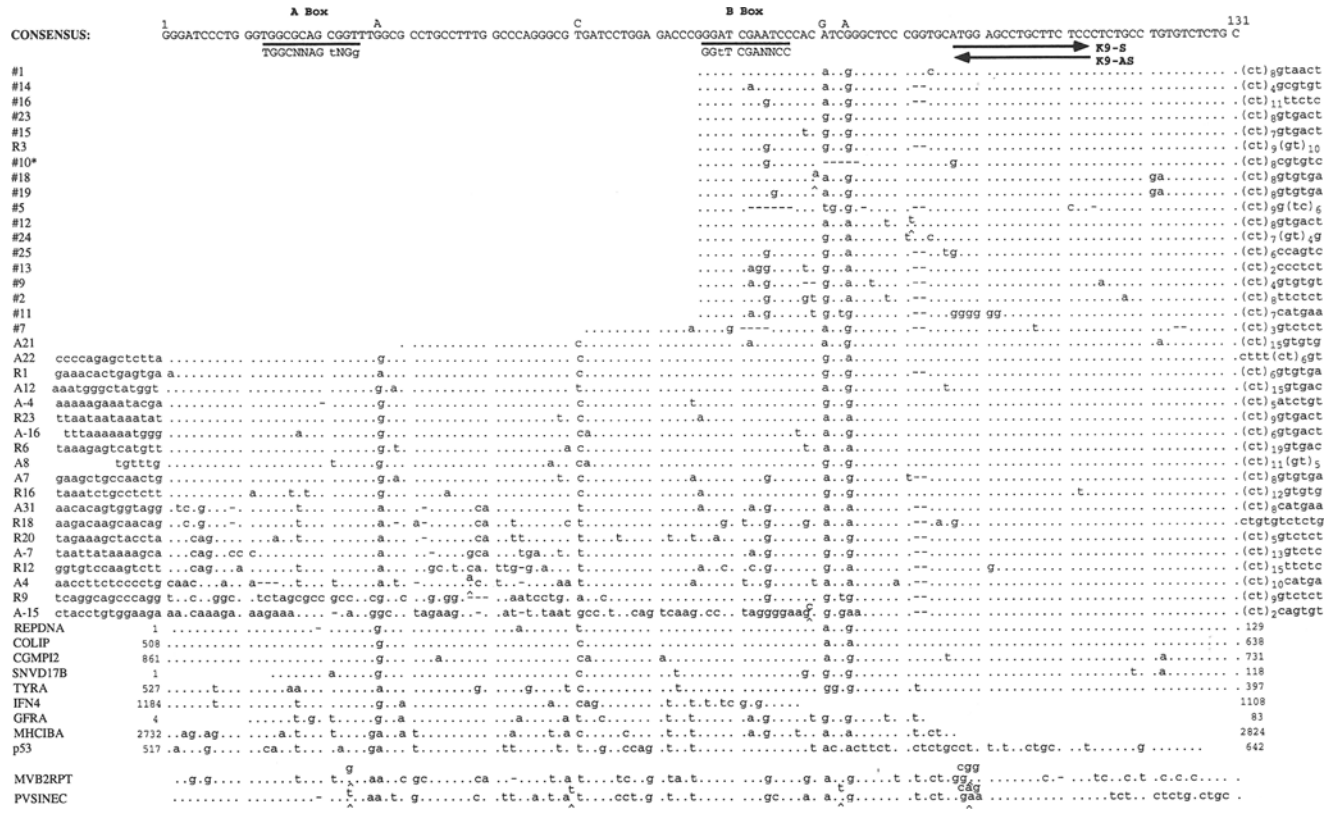

Fig. 1. Sequence comparison of SINEs from *C. familiaris* genomic clones. The top line is the consensus sequence of the canine SINE. Consensus sequence of putative A and B boxes for RNA polymerase III transcriptional control are indicated, with lower case letter indicating divergence from the sequence found in the consensus SINE. Three small insert genomic libraries were screened for canine SINEs as described in the Materials and methods. Clones identified from the *Sau*3AI library were partial clones, since the canine SINE contains 3 *Sau*3AI recognition sites (5′GATC3′). Clone identifiers are presented to the left and indicate those sequences from the *Sau*3AI library (a # sign followed by an number), the Alu I library (an A followed by a number); this library was sampled twice and clones with dashes are those identified in the second sampling, and the *Rsa*I library (an R followed by a number). Lower case letter indicates divergence from the consensus SINE, dots represent sequence identity, dashes represent deletions, and a ∧ sign with the nucleotide immediately above represents an insertion. Clone 10 has an asterisk next to its name to indicate that it was isolated twice from the genomic library. Canine genes from Genbank and EMBL databases found to harbor a SINE and identified by BLAST are illustrated with nucleotides numbers indicating the position of the identified repeat element. Accession numbers and genes for these loci are: REPDNA (emb-X57357)—consensus *C. familiaris* repetitive DNA sequence identified by Minnick and associates (1992) by comparing three elements similar to those identified in this study; COLIP (gb-M63427)—dog pancreatic colipase gene; CGMP12 (emb-Y11309) *C. familiaris* gene encoding cGMP-gated channel alpha subunit; SNVD17B (gb-M73046)—dog inserted sequence in spleen necrosis virus vector provirus clone; TYRA (gb-L47165)—*C. familiaris* tyrosine aminotransferase gene; IFN4 (gb-M28627)—5′ noncoding region of dog interferon-omega; GFRA (gb-L42326)—*C. familiaris* DNA fragment (clone GPCR W); MHCIBA (emb-Z25418)—*C. familiaris* MHC class Ib gene (DLA-79); p53 (gb-U62133)—*C. familiaris* p53 gene. Related SINE from mink (*Mustela vision*) (emb-MVB2RPT) and harbor seal (*Phoca vitulina concolour*) (emb-PVSINEC) are illustrated at the bottom.

to Hybond N+ membrane and probed with radiolabled K9-S oligonucleotide. The use of an oligonucleotide probe, rather than a random primed SINE probe, reduced the possibility that two positively hybridizing genomic fragments of different sizes contained portions of the same repeat and were generated by cleavage within the repeat. The ethidium bromide staining of the *Bam*HI digests is illustrated in Fig. 3A, and the results of probing are presented in Fig. 3B. All clones contained at least one restriction fragment harboring a repetitive element. The differences in intensity among the genomic DNA bands could be interpreted to suggest that some fragments contain more than one SINE. Densitometric scanning of the autorad presented in Fig. 3B allowed us to provide an estimate of the number of repeats within these ten phages. Assigning unity to the less intense peaks (bands), we conservatively estimate the presence of 31 repeats (Fig. 3B and data not shown). Since these repeats would be present over a total distance of ~150 kbp, this suggests an average spacing of ~5 kbp between repeats. Alternatively, hybridization efficiency of oligonucleotide K9-S to a given fragment could be affected by nucleotide mismatches in the target sequence. In this event, it would be difficult to estimate the number of SINEs within a given genomic fragment, simply based on hybridization intensity. A conservative estimate would assume that each positively hybridizing fragment contains only 1 repeat. This would provide a conservative estimate of 18 SINEs over 150 kbp, suggesting an average spacing of ~8.3 kbp. Interval spacings of 5–8.3 kbp suggest that this canid SINE is present at ~360,000–600,000 copies in the diploid canine genome.

We also hybridized λ FIXII plaques with oligonculeotide K9-AS. From a total of 203 clones screened with this probe, 135 plaques were scored as positives (66.5%; data not shown). These results are consistent with the conclusion that these canid SINEs are dispersed, not clustered, given their presence in a large percentage of independent genomic clones. [From this experiment, it is not possible to accurately obtain an estimate of SINE frequency since the number of SINEs per genomic insert cannot be accurately determined, nor is the proportion of recombinants versus non-recombinants in this library known.]

IRS-PCR with Can SINE specific primers. The highly reiterated nature of the canid SINE, as well as the sequence conservation within the downstream half of the repeat, suggested that IRS PCR

R1 GTTAATAAGAAACACTGAGTGAA---------(CT)$_6$GTGTGACTATCAT(A)$_{11}$TT(A)$_{12}$GAAACACTGAGTGCAGGGGAC

A12 CCTTAGAAATGGGCTATGGT---------(CT)$_{15}$GTGACTATCATAAATAAAT(A)$_5$TT(A)$_{17}$GAAATGGGCTATGGTGTGTG

R23 TGAGTTTAATAATATAAATAT---------(CT)$_9$GTGACTATCATAAATAAAT(A)$_5$TTT(A)$_{11}$TTTAATAATATAAATATAAA

R6 CTTATTTAAAGAGTCATGTT---------(CT)$_{19}$GTGACTATCATAAATAAAT(A)$_{10}$T(A)$_6$AAAAAGTCATGTTGCCTCTGT

R16 CTTCACAATAAATCTGCCTCTT---------(CT)$_{12}$(GT)$_9$GACTGTCATAAAT(A)$_7$TAAATT(A)$_9$TAAATCTGCCTCTTGACAAGCT

A31 GGAACACAGTGGTAGG---------(CT)$_8$CATGAATAAATAAAATCTT(A)$_{10}$GTGGTAGGATGCCTGG

R18 AAGACAAGCAACAG---------(CT)$_8$GTGTCTCTCATGAATAAATAAATAAAATCTT(A)$_{12}$(GA)$_{11}$CAAGCAACAAATAAGAG

R20 TATAGAAAGCTACCTA---------(CT)$_5$GTCTCTCATGGATAAATAAATAAATCTT(A)$_{13}$GG(A)$_8$TAAAACTACCTAATGTGTG

A-7 GGGAACTTAATTATAAAAGCA----------(CT)$_{13}$GTCTCTATGAATAAAT(A)$_9$TCTTTAT(A)$_7$GAATTATAAAAACATGTTAA

R9 AACAAAAGGAAAGTATCAGGCAGCCCAGGT---------(CT)$_9$GTCTCTCATTAATAAATAAAATCTTT(A)$_{17}$GAAAGTATCTTGAAA

COLIP ACTTAGAACAGATAACTT---------(CT)$_7$GTGACTATCATGAATAAATAAATAAAATCTT(A)$_7$TATT(A)$_{11}$GAACAGATAACTTGTC

CGMPI2 ATTTTTTAAAATCAATTAATTG--------(CT)$_7$GAGACTATCATAAATAAATAAAAATTG(A)$_9$TTAAAATCAATTAATTAACAT

TYRA AATTAGAAGCAGATGGA--------CTCTTC(CT)$_4$GTGTGTGACTATTATAAATAAATAAAAATAATT(A)$_{12}$GAAGCAGATGGAACTT

P53 CTTGGCTTTTAAAGTAAGGTTCTGG--------AAAATAAAGTAAGCTTCTGGAATAGACC

Fig. 2. Comparison of canid SINE flanking sequences. The flanking sequences of 14 canid SINE family members are compared. Direct repeat sequences are in bold and underlined. The region corresponding to the consensus core SINE sequence is represented by a dashed line.

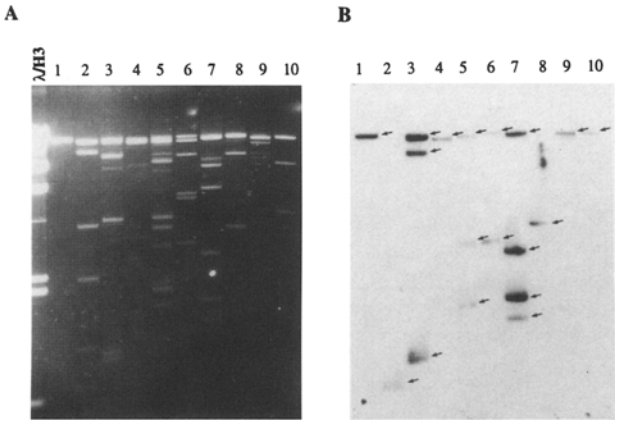

Fig. 3. Repetition frequency of the canid SINE. Hybridization profile of DNA from ten canine genomic phages digested with *Bam*HI and probed with radiolabeled K9-S oligonucleotide. Following fractionation on a 0.8% agarose gel, DNA was transferred onto Hybond-N+ and processed as described in the Materials and methods. Phage numbers are indicated above the panel, and positively hybridizing genomic DNA fragments are indicated with arrows. Molecular weight markers are from λ digested with *Hin*dIII. **(A)** Ethidium bromide staining of gel and **(B)** Southern blot profile after hybridization with radiolabeled K9-S oligonucleotide. Densitometric scanning of the autoradiograph was performed on an LKB 2202 UltroScan Laser Densitometer coupled to an LKB 2220 Recording Integrator.

was possible in the dog with oligonculeotides to the conserved region of the SINE. Amplification of genomic DNA with the K9-S oligonucleotide indicated that this primer was canine specific, since no amplification products were observed with human or mouse genomic DNA (Fig. 4A, lanes 2 and 3), whereas a smear of products was clearly visible when DNA from MDCK cells or from a Border Collie was amplified (lanes 4 and 5). Similar results were obtained when the K9-AS oligonucleotide was used in an IRS PCR (data not shown).

To determine whether the *Can* SINEs we had isolated were present in a variety of related Canidae, we performed slot blot analysis on DNA of representative species from the Canidae family (dog [*Canis familiaris*], gray wolf [*Canis lupus*], arctic fox [*Alopex lagopus*], side-striped jackal [*Canis adustus*], and black-backed jackal [*Canis mesomelas*] Fig. 4B). In addition, DNA from human, mice, and cat were also analyzed by hybridization. Canid DNA hybridized to radiolabeled K9-S oligonucleotide probe, whereas DNA from human, mouse, or cat did not (Fig. 4B). These experiments demonstrate that this canid SINE is present in closely related wolf and jackal, as well as the more distantly related arctic fox. Consistent with these results, IRS-PCR performed with fox, jackal, and wolf DNA with oligonucleotide K9-S produced a large number of products, visualized as a smear on an agarose gel (Fig. 4C). The reason for the differences in product size between the jackal samples (lanes 5 and 6) and the other canid samples (lanes 2–4) is not known, but may be related to the fact that both sets of jackal DNA were extensively degraded, relative to the other samples (J. Pelletier; data not shown). These results attest to the highly specific nature of inter-SINE amplification in canids.

IRS PCR products are polymorphic. To determine whether IRS PCR products were polymorphic, we isolated five of these products and determined their nucleotide sequence. One clone was found to be polymorphic and is presented in Fig. 5A. Isolation and sequencing of several clones of this product revealed five sequence differences between allelic variants of this product (Fig. 5A). An oligonucleotide was designed to the unique middle portion of the IRS product and allowed us to sample a complex pool of IRS-PCR products by hybridization to total IRS products that had been fractionated on SSCP gels (Fig. 5B). The results indicate that the IRS product under analysis is polymorphic between different dog breeds (for example, compare lane 2 with lane 3), as well as within a given breed (for example, compare lane 2 with lane 3 and lane 14 with 15). By SSCP analysis of the two allelic variants of this locus presented in Fig. 5A, we have deduced that the *b* allele corresponds to the sequence presented, whereas the *a* allele harbors the five sequence changes indicated in Fig. 5A (M. Das and J. Pelletier; data not shown). Our results suggest that some IRS-PCR products will be sufficiently polymorphic to allow filter-based genotyping to be applied to the dog.

Discussion

We have characterized a family of highly repeated dispersed DNA sequences present in the *C. familiaris* genome. Three examples of this family from dogs had previously been reported by Minnick and associates (1992), and our work extends the analysis of this family, particularly in terms of family size and individual member heterogeneity. Like other SINEs, members of the *Can* SINE family can be seen as closely related families of pseudogenes amplified through RNA intermediates from a number of conserved and tran-

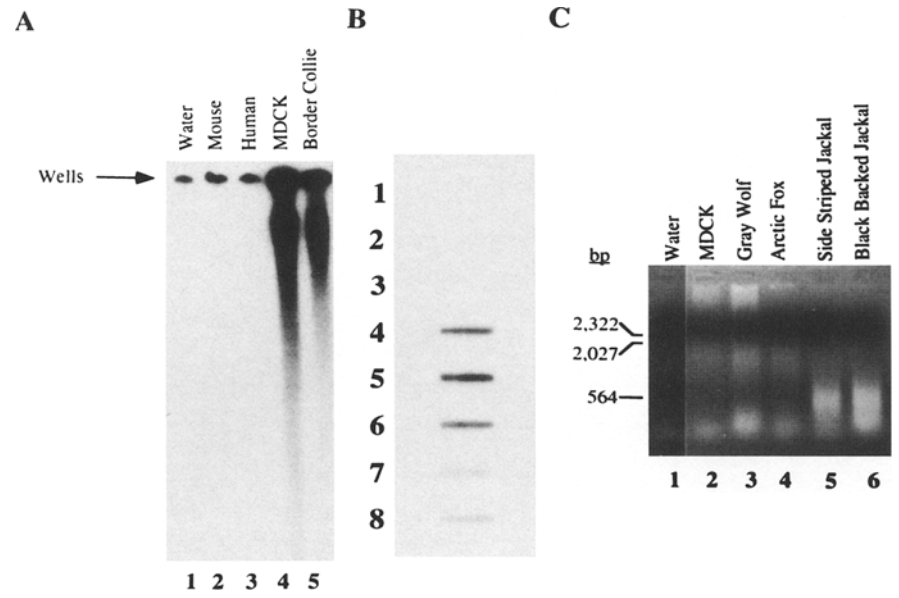

Fig. 4. IRS PCR from canine genomic DNA. **(A)** IRS PCR with oligonucleotide K9-S. Oligonucleotide K9-S was radiolabeled with T4 PNK and λ-[^{32}P]-ATP and used in a PCR as described in the Materials and methods. Amplification products (10% of the reaction; 2 μl) were loaded onto a 6% nondenaturing polyacrylamide gel. The origin of input genomic DNA in the PCR is indicated above the well, and the sizes of the PCR products visualized in lanes 4 and 5 are >1000 bp. **(B)** Slot blot analysis showing the presence of the *Can* SINE in several species from the Canidae family. Approximately, 5 μg of each purified DNA was denatured in 100 μl 0.4 *N* NaOH/0.6M NaCl for 20 min, then slot blotted onto Hybond N+ membrane. Hybridizations were performed with oligonucleotide K9-S (10^7 cpm/ml), and washes were for 40 min at room temperature in 2 × SSC/0.05% SDS and then 0.1 × SSC/0.01% SDS. DNA samples: 1, human; 2, mouse; 3, cat; 4, MDCK cells (*C. familiaris*); 5, gray wolf (*C. lupus*); 6, arctic fox (*Alopex lagopus*); 7, side-striped jackal (*C. adustrus*); 8, black-backed jackal (*C. mesomelas*). Slot blot was exposed to X-Omat (Kodak) film for 6 h at room temperature. **(C)** IRS PCR in several species of the Canidae family. Amplification reactions were performed from genomic DNA that had been isolated from tissue obtained from the sources indicated above the panel. Molecular weight markers are from λ digested with *Hin*dIII. A smear of products >1000 bp was obtained from MDCK (lane 2), gray wolf (lane 3), and arctic fox (lane 4) DNA; whereas a set of smaller products were obtained from side striped (lane 5) and black-backed (lane 6) jackal DNA. The reason for the difference in product size between jackal and other canid samples is not known, but may relate to the fact that both sets of jackal DNA were extensively degraded.

scriptionally active master genes that appeared at different time periods during evolution.

Can SINEs have consensus internal Pol III control regions (Fig. 1) that will function in vitro (Fuhrman et al. 1981), suggesting that they may be capable of coding for Pol III transcripts. Sequence studies of genomic copies of the *Can* SINEs revealed they are flanked on either side by a direct repeat of 8 to 17 base pairs (Fig. 2), similar to what has been described for many other dispersed repetitive DNA families (for a review, see Hastie 1989). The presence of these repeats suggests duplication of preexisting genomic insertion sites, making it likely that members of this family have arise by retrotransposition. The presence of an A-rich region of variable length in the *Can* SINE is a feature also present in other SINEs and is thought to reflect the ability of the RNA to fold over on itself and act as a primer for reverse transcription. Screening of a canine liver cDNA library with oligonucleotide K9-S indicated that the canid SINE (or at least the downstream segment) is also present in some mRNA transcripts (J. Pelletier and L.L. Chu; data not shown).

We estimate that members of this particular family are present at least once every ~5–8.3 kbp, constituting ~360,000–600,000 copies. Since each repeat is ~150 nucleotides in length [including the $(CT)_n$ repeat and poly A tail], this family represents ca. 1.8–3% of the dog genome (1.5×10^9 bp per haploid genome)—a frequency lower than Alu repeats in humans (5–6% of the genome) and slightly higher than B1 repeats in mice (0.7–1% (Hastie 1989).

Our results, and those of other groups (Minnick et al. 1992; Coltman and Wright 1994), indicate that members of the *Can* SINE family are also present in wolf, jackal, and the more distantly related fox (Fig. 4). Members of this repeat family have also been identified in mink (family Mustelidae; Lavrentiea et al. 1989) and seal (Phocidae) genomes (Coltman and Wright 1994). The *Can* SINE shows an average sequence similarity of ~70% to the seal and mink SINEs (Fig. 1; Minnick et al. 1992; Coltman and Wright 1994). *Can* SINEs were not detected in the cat genome (superfamily Feloidea; Minnick et al. 1992; Coltman and Wright 1994; this report). Results by Minnick and colleagues (1992) indicated that the *Can* SINE is also present in the family Canidae (gray fox), but not in the family Ursidae (black bear), Procyonidea (raccoon), or Mustelidae (ferret) of the order Carnivora, nor in the superfamily Feloidea. The reason for the discrepencies between the findings of Minnick and associates (1992) and Coltman and Wright (1994) with respect to the absence or presence of this repeat in the family Mustelidae is not known. Recently, alignment of a small number of canine, mink, and seal SINEs allowed van der Vlugt and Lenstra (1995) to design oligonucleotides to various domains of the *Can* SINE and to undertake IRS PCR. From our primary sequence analysis of several members of the *Can* SINE family (Fig. 1), their oligonucleotide design may not have been optimal for IRS amplification in the dog. Their finding of *Can* SINEs in cat DNA by IRS PCR is contrary to genomic hybridization data (Minnick et al. 1992; Coltman and Wright 1994; this study). Given the absence of *Can* SINEs in human or rodent genomes, these elements should provide useful tags for detecting Canoidea-specific DNA in somatic cell genetics or molecular genetics.

We have demonstrated that portions of the canid SINE sequence are sufficiently conserved to allow IRS PCR from dog DNA (Fig. 4). Polymorphisms from these PCR products can be detected by SSCP fractionation, indicating that these markers could be developed for filter-based genotyping in the dog (Fig. 5). Allelic variation between inter-Alu PCR products has been used in humans to determine linkage of pseudovitamin D-deficiency rickets to Chr 12 (Zietkiewicz et al. 1992). Polymorphisms in IRS products may arise from (i) length variability and/or (ii) sequence variability in the intervening sequence. Additionally, the presence or absence of a specific IRS product among inbred mouse strains is a powerful mapping tool in mouse genomics (Hunter et al 1996). Although the probe we have characterized proved to be polymorphic, not all IRS-products analyzed showed allelic variation (M. Das and J. Pelletier; data not shown). Digesting the target DNA to generate smaller target products may reveal polymorphisms not detectable on larger targets, given the size limitations of SSCP analysis (Liu and Sommer, 1994). The generation of a collection of well-mapped IRS products would be invaluable to many facets of canine genomics.

Acknowledgments. This work is supported by grants from the National Cancer Institute and Medical Research Council of Canada to J. Pelletier, who is an MRC Scientist.

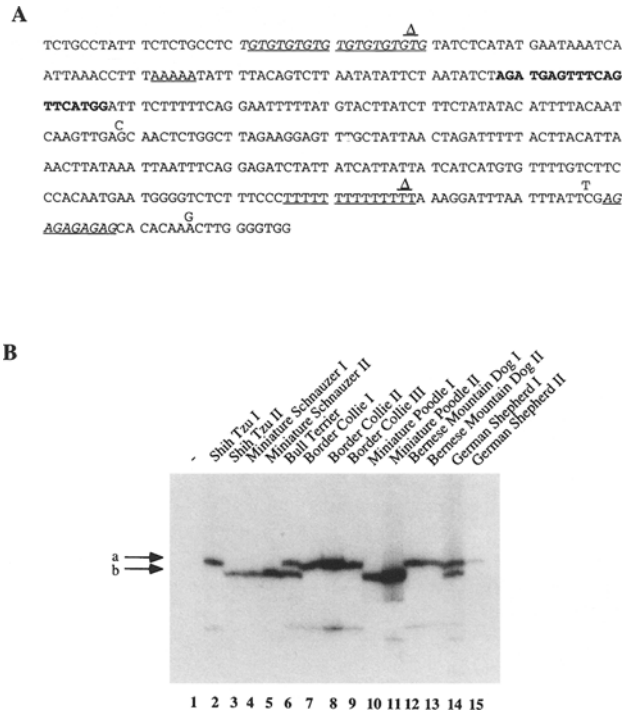

Fig. 5. IRS PCR products are polymorphic. **(A)** Nucleotide sequence of an IRS product isolated from a PCR with oligonucleotide K9-S and genomic DNA from a German Shepherd. The sequence of the K9-S is not shown but occurs immediately upstream and downstream of the sequence shown. GT/GA tracts are underlined and in italics, whereas A/T-rich regions are underlined. (Note that this IRS PCR product does not have a characteristic $(CT)_n$ tract.) The sequence chosen for oligonucleotide design is in bold. Nucleotide differences between two isolated clones of this IRS-PCR product are illustrated, with differences indicated above the sequence. A △ symbol indicates a deletion of 2 bp. The sequence shown corresponds to the *b* allele in (B), whereas the product with the five sequence differences indicated corresponds to the *a* allele. **(B)** Southern blot with the IRS oligonucleotide derived from (A) against total IRS products that had been fractionated on an SSCP gel and transferred to Nylon Hybond N+. The breed of origin of each IRS product is shown above each lane, and numbers indicate different animals from each breed.

References

Benton WD, Davis RW (1977) Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196, 180–182

Brooks-Wilson A, Smailus DE, Weier H-UG, Goodfellow PJ (1992) Human repeat element-mediated PCR: cloning and mapping of chromosome 10 DNA markers. Genomics 13, 409–414

Chumakov IM, Le Gall I, Billault A, Ougen P, Soularue P, Guillou S, Rigault P, Bui H, de Tand M-F, Barillot E, Abderrahim H, Cherif D, Berger R, Le Paslier D, Cohen D. (1992) Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. Nature Genet 1, 222–225

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

Deininger PL (1989) SINEs: short interspersed repeated DNA elements in higher eukaryotes. In *Mobile DNA* DE Berg, MM Howe, eds. (Washington, DC: Am Soc Microbiol), pp 619–636

Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW (1981) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J Mol Biol 151, 17–33

Fischer PE, Holmes NG, Dickens HF, Thomas R, Binns MM, Nacheva EP (1996) The application of FISH techniques for physical mapping in the dog (*Canis familiaris*). Mamm Genome 7, 37–41

Francisco LV, Langston AA, Mellersh CS, Neal CL, Ostrander EA (1996) A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. Mamm Genome 7, 359–362

Fuhrman SA, Deininger PL, La Porte P, Friedmann T, Geiduschek EP (1981) Analysis of transcription of the human Alu family ubiquitous repeating element by eukaryotic RNA polymerase III. Nucleic Acids Res 9, 6439–6456

Guzzetta V, Montes de Oca-Luna R, Lupski JR, Patel PI (1991) Isolation of region-specific and polymorphic markers from chromosome 17 by restricted Alu polymerase chain reaction. Genomics 9, 31–36

Hastie ND (1989) Highly repeated DNA families in the genome of *Mus musculus*. In *Genetic Variants and Strains of the Laboratory Mouse,* MF Lyon, AG Searle, eds., (London/New York: Oxford Univ. Press), pp 559–573

Hunter KW, Ontiveros SD, Watson ML, Stanton Jr VP, Gutierrez P, Bhat D, Rochelle J, Graw S, Ton C, Schalling M, Aburatani H, Brown SDM, Seldin MF, Housman DE (1994) Rapid and efficient construction of yeast artificial chromosome contigs in the mouse genome with interspersed repetitive sequence PCR (IRS-PCR): Generation of a 5-cM, >5 megabase contig on mouse chromosome 1. Mamm Genome 5, 597–607

Jelinek WR (1982) Repetitive sequences in eukaryotic DNA and their expression. Annu Rev Biochem 51, 813–844

Jurka J, Smith T (1988) A fundamental division in the Alu family of repeated sequences. Proc Natl Acad Sci USA 85, 4775–4778

Lavrentieva MV, Rivkin MI, Shilov AG, Kobetz ML, Rogozin IB, Serov OL (1989) B2-like repetitive sequence of American mink genome. Bokl Akad Nauk SSSR 307, 226–230

Ledbetter SA, Nelson DL, Warren ST, Ledbetter DH (1990) Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. Genomics 6, 475–481

Liu J, Stanton Jr VP, Fujiwara TM, Wang J-X, Rezonkew R, Crumley MJ, Morgan K, Gros P, Housman D, Schurr E (1995) Large-scale cloning of human chromosome 2-specific yeast artificial chromosomes (YACs) using an interspersed repetitive sequences (IRS)-PCR approach. Genomics 26, 178–191

Liu Q, Sommer SS (1994) Parameters affecting the sensitivities of dideoxy fingerprinting and SSCP. PCR Methods Appl 4, 97–108

McCarthy L, Hunter K, Schalkwyk L, Riba L, Anson S, Mott R, Newell W, Bruley C, Bar I, Ramu E, Housman D, Cox R, Lehrach H (1995) Efficient high-resolution genetic mapping of mouse interspersed repetitive sequence PCR products, toward integrated genetic and physical mapping of the mouse genome. Proc Natl Acad Sci USA 92, 5302–5306

Minnick MF, Stillwell LC, Heineman JM, Stiegler GL (1992) A highly repetitive DNA sequence possibly unique to canids. Gene 110, 235–238

Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, Caskey CT (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc Natl Acad Sci USA 86, 6686–6690

Ostrander EA, Sprague Jr GF, Rine J (1993) Identification and characterization of dinucleotide repeat $(CA)_n$ markers for genetic mapping in the dog. Genomics 16, 207–213

Ostrander EA, Mapa F, Yee M, Rine J (1995) One hundred new simple sequence repeat-based markers for the canine genome. Mamm Genome 6, 192–195

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual,* 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press)

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74, 5463–5467

van der Vlugt HHK, Lenstra JA (1995) SINE elements of carnivores. Mamm Genome 6, 49–51

Yuzbasiyan-Gurkan V, Blanton SH, Cao Y, Ferguson P, Li J, Venta PJ, Brewer GJ (1997) Linkage of a microsatellite marker to the canine copper toxicosis locus in Bedlington Terriers. Am J Vet Res 58, 23–27

Zietkiewicz E, Labuda M, Sinnett D, Glorieux FH, Labuda D (1992) Linkage mapping by simultaneous screening of multiple polymorphic loci using *Alu* oligonucleotide-directed PCR. Proc Natl Acad Sci USA 89, 8448–8451

Zucman J, Delattre O, Desmaze C, Azambuja C, Rouleau G, De Jong P, Aurias A, Thomas G (1992) Rapid isolation of cosmids from defined subregions by differential *Alu*-PCR hybridization on chromosome 22-specific library. Genomics 13, 395–401.