2726–2730 Nucleic Acids Research, 1994, Vol. 22, No. 14

Can SINEs: a family of tRNA-derived retroposons specific to the superfamily Canoidea

David W.Coltman and Jonathan M.Wright*
Department of Biology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

Received May 13, 1994; Accepted June 8, 1994 EMBL accession nos: Z33495–Z33501 (incl.)

ABSTRACT

A repetitive element of approximately 200 bp was cloned from harbour seal (*Phoca vitulina concolour*) genomic DNA. The sequence of the element revealed putative RNA polymerase III control boxes, a poly A tail and direct terminal repeats characteristic of SINEs. Sequence and secondary structural similarities suggest that the SINE is derived from a tRNA, possibly tRNA-alanine. Southern blot analysis indicated that the element is predominately dispersed in unique regions of the seal genome, but may also be present in other repetitive sequences, such as tandemly arrayed satellite DNA. Based on slot-blot hybridization analysis, we estimate that 1.3×10^6 copies of the SINE are present in the harbour seal genome; SINE copy number based on the number of clones isolated from a size-selected library, however, is an order of magnitude lower ($1-3 \times 10^5$ copies), an estimate consistent with the abundance of SINEs in other mammalian genomes. Database searches found similar sequences have been isolated from dog (*Canis familiaris*) and mink (*Mustela vison*). These, and the seal SINE sequences are characterized by an internal CT dinucleotide microsatellite in the tRNA-unrelated region. Hybridization of genomic DNA from representative species of a wide range of mammalian orders to an oligonucleotide (30mer) probe complementary to a conserved region of the SINE confirmed that the element is unique to carnivores of the superfamily Canoidea.

INTRODUCTION

Eukaryote genomes are composed mainly of non-coding and frequently repetitive DNA. Repetitive DNA is commonly classified as being either tandemly arrayed, (e.g. satellite, minisatellite and microsatellite DNAs) (1), or dispersed (2). Homogenization and propagation of tandem repeats in the genome is thought to proceed by several mechanisms such as unequal crossing-over, slippage during replication or saltatory amplification (reviewed in 1, 3, 4) whereas dispersed repeats, in many instances, appear to be propogated by retroposition via an RNA intermediate (5, 6). Dispersed repeats are classified as (i) short, interspersed elements (SINEs), which are less than 500 bp and have characteristics of RNA polymerase III transcripts (5–7), and (ii) long interspersed elements (LINEs) which can be several kilobases long and are probably derived from RNA polymerase II transcripts (8, 9). Retroposition of SINEs is thought to involve transcription of an active copy of a SINE gene by RNA polymerase III, reverse transcription of the RNA in the nucleus, and integration of the cDNA at a new genomic site, probably a nicked, A-rich region (5, 6, 10).

SINEs were first isolated from mammals, with distinct SINE families present in different lineages (11). More recently, SINE families have been found in other eukaryotes such as fishes (12), higher plants (13) and cephalopods (14). Many SINE families have a composite structure that includes a tRNA-like region at the 5′ end which tends to be conserved between families, a central family-specific or tRNA-unrelated region, and an AT rich region at the 3′ end (12–14). During the isolation of microsatellites from a harbour seal (*Phoca vitulina concolour*) genomic library, we cloned several SINE sequences. Based on hybridization analysis, their distribution appears restricted to canoid carnivores.

MATERIALS AND METHODS

Library construction and screening

High molecular weight genomic DNA of a single harbour seal was extracted from blood by standard methods (16) and digested with a mixture of the restriction enzymes *Alu*I, *Hinc*II and *Hae*III (Pharmacia). Fragments were separated in a 1% low melting point agarose gel and DNA from a gel slice containing the 300–800bp size fraction was extracted by standard techniques (16). DNA was ligated into *Sma*I digested dephosphorylated pUC18 (Pharmacia). 'Max-efficiency' DH5α cells (BRL) were transformed, plated on selective media and colonies were immobilized on Amersham N+ nylon membranes. Membranes were prehybridized for 3 hours in 5×SSPE / 0.1% SDS / 5×Denhardt's / 10 μg/mL RNA at 58°C. For a probe, 100 ng of $(CT)_{15}$ oligonucleotide was end-labelled with T4 polynucleotide kinase (Pharmacia) and γ-[^{32}P]-dATP to high specific activity. The probe was added to the mixture and hybridization was allowed to proceed overnight at 58°C. Membranes were washed twice at room temperature in 2×SSPE / 0.1% SDS for 15 minutes and exposed to X-ray film overnight at −70°C. Positively hybridizing colonies were picked, cultured

*To whom correspondence should be addressed

and stored as stocks at −70°C. Plasmid DNA was prepared for sequencing using the protocol of Goode and Feinstein (17). Sequencing was performed using T7 DNA polymerase (Pharmacia) and forward and reverse universal primers.

Southern blot analysis

Ten micrograms of harbour seal genomic DNA was digested with *Eco*RI, *Hinf*I, *Hinc*II, *Hae*III or *Alu*I according to the manufacturer's instructions (Pharmacia). Digests were loaded in a 1% agarose gel and separated by electrophoresis at 1.5 V/cm for 24 hours. DNA was transferred and fixed to Amersham N+ nylon membranes by standard capillary blotting techniques (16). The membrane was prehybridized for 3 hours at 65°C in 5×SSPE / 0.1% SDS / 5×Denhardt's / 10 μg/mL RNA. 100 ng of 30mer oligonucleotide, complementary to a conserved region in canoid SINEs (see Figure 1B), was end-labelled to high specific activity and used as a probe. Hybridization proceeded overnight at 65°C. The membrane was washed twice at room temperature and once at 42°C in 2×SSPE / 0.1% SDS, and exposed to X-ray film for 36 hours at −70°C with intensifying screens.

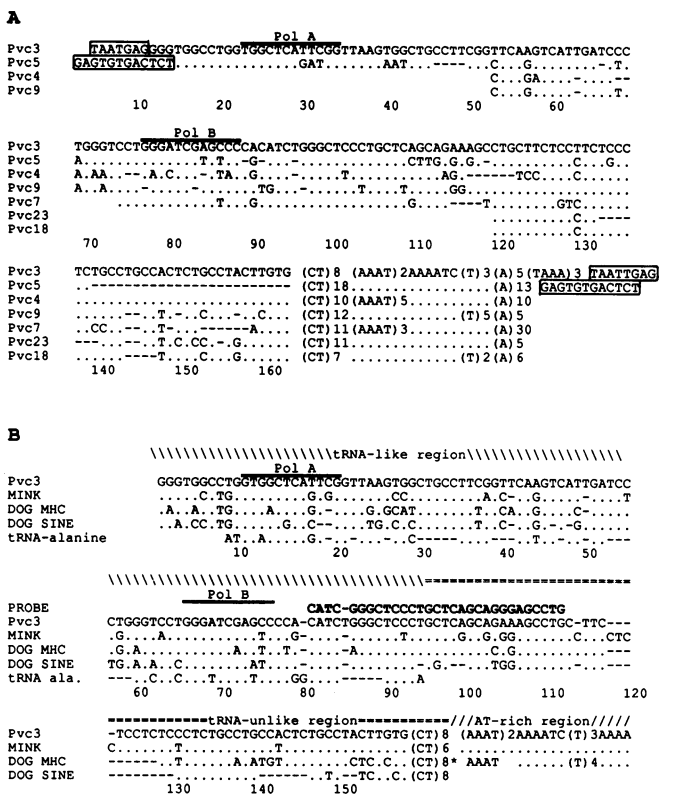

Figure 1. Nucleotide sequence of *Can* SINEs. Panel **A** shows the alignment of two complete and five partial SINEs isolated from harbour seal. Terminal direct repeats are boxed. Panel **B** shows the alignment of one complete harbour seal SINE, Pvc3, with SINEs previously isolated from mink and dog (GenBank accession numbers X52381, Z25418, and X57357) and a mouse tRNA-alanine. Putative tRNA-related, tRNA-unrelated, and AT-rich regions in *Can* SINEs are highlighted, and RNA polymerase III control boxes A and B are overlined. The oligonucletide used in hybridization studies is shown aligned to the Pvc3 sequence in panel B. In both panels, the symbols (.) and (−) indicate nucleotide identity and gaps inserted to maximize sequence similarity, respectively. The CT repeats are shown as perfect for the sake of clarity, however, some are imperfect containing single base substitutions. The symbol (*) in the dog MHC intron indicates a 13 base pair insertion (TTCATGAATACAT).

Slot blot analysis

Slot blots were used to estimate copy number and to assay for the presence of homologous SINEs in other mammalian orders. To estimate copy number, duplicate serial dilutions of plasmid DNA containing a complete SINE insert and of harbour seal genomic DNA were applied to an Amersham N+ nylon membrane. The 'zoo blot' consisted of 2 μg of genomic DNA from a fish (Atlantic salmon [*Salmo salar*]) and representatives of the mammalian orders Carnivora (superfamily Feloidea-domestic cat [*Felis catus*]; superfamily Canoidea-harbour seal, domestic dog [*Canis familiaris*], grey wolf [*Canis lupus*], coyote [*Canis latrans*], and American mink [*Mustela vison*]), Lagomorpha (rabbit [*Sylvilagus*]), Artiodactyla (deer [*Odocoileus virginianus*], cow [*Bovis taurus*]), Rodentia (mouse [*Mus musculus*]), Primates (human [*Homo sapiens*]) and Cetacea (sperm whale [*Physeter macrocephalus*]) applied to an Amersham N+ membrane. DNA was denatured in 100 μl of 0.4M NaOH/ 0.6M NaCl for 5 minutes, applied to the membrane, fixed by baking for 2 hours at 80°C, then hybridized to the oligonucleotide derived from the SINE sequences shown in Figure 1B. To estimate copy number, portions of the membrane containing plasmid and genomic dilutions was cut into equal area and the radioactivity measured by liquid scintillation spectrometry. The mean (n=2) amount of radioactivity (corrected for background) for each dilution was plotted versus the amount of DNA applied, and copy number estimated by comparing plasmid and genomic DNA applied at equivalent hybridization signal (data not shown).

RESULTS AND DISCUSSION

Sequence analysis

Sixteen clones out of 400 recombinant colonies hybridized to the $(CT)_{15}$ oligonucleotide probe. Twelve clones had CT repeats embedded in a SINE-like sequence. The sequences of 7 of these clones which contained at least 20 bases of the SINE sequence upstream from the CT microsatellite were aligned by eye and gaps introduced to maximize sequence similarity (Figure 1A). The other 5 clones containing SINE-like sequences had been cleaved internally close to the poly-A region by one of the restriction enzymes used during construction of the genomic library (data not shown). Two clones contained the complete SINE sequence, as evidenced by the presence of direct terminal repeats (Figure 1A). The presence of sequences similar to RNA polymerase III promoter binding sites, the pol III A and pol III B boxes near the 5′ end, a poly A region at the 3′ end of the sequence, and direct terminal repeats are characteristic of SINE elements (5,6). The putative pol III A and B boxes are consistent in sequence composition and relative position within the element to pol III A and B boxes found in other SINEs and tRNAs (6). The average sequence similarity between the harbour seal SINEs (excluding the highly variable CT microsatellite-poly A region) is 71.8%. Most mammalian SINEs have an A-tract at their 3′ end, occasionally punctuated by other bases. The presence of these A-tracts has led to speculation that they arise by an aberrant polyadenylation process of the RNA molecules during retroposition. The presence of AAAT stretches and the CT repeat at the 3′ end of seal SINEs is unusual and suggests that the A-rich sequence of the retroposed seal SINE may have been encoded in the source or master gene (10).

Molecular sequence database searches conducted using BLAST (18) found similar SINE sequences in mink (a 'b-2 like repeat') (19) and dog (a SINE element reported by Minnick *et al.*, and an intron in a dog MHC gene) (20,21). An alignment of these sequences with the sequence of a complete seal SINE, clone Pvc 3, is shown in Figure 1B. The average sequence similarity of these other SINEs to the seal SINE is 70.6%. We have identified more plausible pol III A and B boxes in the dog SINE than that proposed by Minnick *et al.* (20).

We compared the consensus sequence of the dog, mink, and the 2 complete seal SINEs to mammalian tRNA sequences listed in GenBank and in the literature to determine the location of tRNA-related and tRNA-unrelated regions of the SINE (data not shown). While most tRNA genes contain regions of homology in the pol III A and B promoter regions with overall sequence similarities of 40−60%, the SINE consensus sequence exhibits the greatest similarity (~69%) to a mouse tRNA-alanine (22). Furthermore, sequence similarity with tRNA-alanine is distributed more evenly throughout the length of the tRNA molecule than it is with other mammalian tRNAs (see Figure 1B). Figure 2 illustrates a possible cloverleaf secondary structure for the tRNA-related region of the consensus canoid SINE sequence. This structure demonstrates conservative base-pairing of the putative pol III A and B regions in the D and T loop stems similar to that observed in tRNA-alanine (23).

The seal, mink and dog SINEs contain a microsatellite dinucleotide CT repeat in the tRNA-unrelated region which suggests that they may provide a general source of genetic markers for genome mapping, pedigree or population studies. Preliminary experiments using the polymerase chain reaction and primers complementary to flanking sequences have shown that a single SINE locus (Pvc 5) can be reliably amplified, however, this locus appears to be monomorphic in a few individuals from the population studied (data not shown).

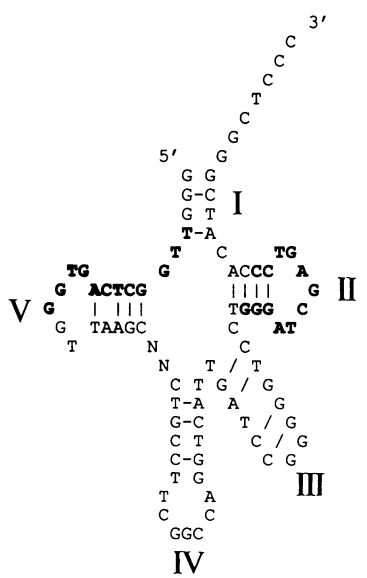

Figure 2. Possible secondary structure of *Can* SINEs. The consensus sequence is shown as DNA with regions of base-pairing shown by lines. The 5′ and 3′ termini, aminoacyl stem (I), the dihydrouridine loop (II), the variable loop (III), the anticodon loop (IV) and the pseudouridine loop (V) are indicated. The putative RNA polymerase III promoter regions are shown in bold type.

The sequences of canoidean SINEs are generally dissimilar to the tRNA-unrelated regions of other well-characterized mammalian SINE families such as the *Alu* sequence of primates and the B1, B2 and ID sequence of rodents, (data not shown). Coupled with the sequence similarity of seal, dog and mink SINEs (>70%), this implies a common evolutionary origin within the order Carnivora. To test this hypothesis and to determine the genomic organization of these elements in the harbour seal, we

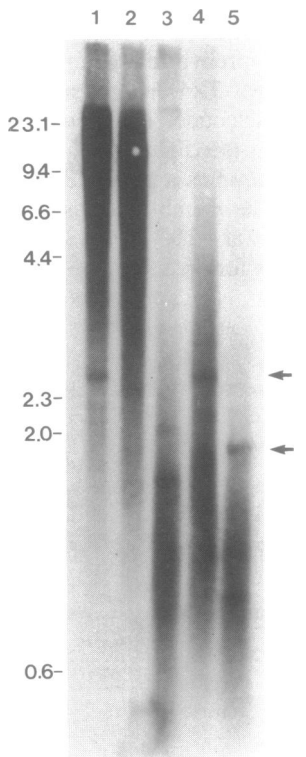

Figure 3. Genomic organization of *Can* SINEs in the harbour seal. The Southern blot of harbour seal genomic DNA digested with *Eco*RI, *Hinf*I, *Hae*III, *Hinc*II and *Alu*I (Lanes 1−5, respectively) was hybridized to an oligonucleotide derived from the *Can* SINE sequence. Size standards (in kilobase pairs) shown at left are derived from lambda phage DNA digested with *Hin*dIII (Pharmacia). Arrows indicate the position of highly repeated sequences containing SINEs.

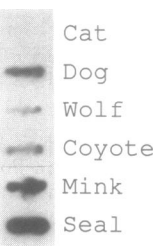

Figure 4. Distribution of *Can* SINEs in the Carnivora. Slot-blot showing hybridization of the oligonucleotide derived from the *Can* SINE sequence to genomic DNA from representatives of the superfamily Canoidea (dog, wolf, coyote, mink and seal), and its lack of hybridization to the representative of the superfamily Feloidea (cat).

designed an oligonucleotide complementary to a conserved 30 base pair region straddling the tRNA-like and family specific regions of the SINE for use as a probe in hybridization analysis (Figure 1B). Although high stringency conditions were used, as a precaution the oligonucleotide was designed to avoid hybridization to tRNA genes, unrelated CT repeats and poly A regions in the genome which might occur if an entire SINE was used as a hybridization probe.

Genomic organization of canoid SINES in the harbour seal

Hybridization of the synthetic oligonucleotide to numerous restriction fragments of various molecular weight observed in the Southern blot of harbour seal DNA show that the element is predominately dispersed in the harbour seal genome (Figure 3). However, the presence of one or two intense bands in each digest indicates a possible association of the SINE with larger repeats. None of the harbour seal SINE sequences possess an *Eco*RI recognition sequence, yet there is a discrete band present at 2.5 kb in the *Eco*RI digest (Figure 3). An equally intense band of the same size (2.5 kb) is also present in the *Hinc*II digest, and at 1.9 kb in the *Alu*I digest. These restriction fragments correspond to some of the bands seen in ethidium bromide stained restriction digests of harbour seal genomic DNA (data not shown). This result suggests that some of the harbour seal SINEs may have integrated into other types of repetitive sequences, such as a LINE or a monomer unit of a satellite DNA array, prior to their amplification.

By quantitative slot-blot hybridization of the oligonucleotide probe to SINE-containing plasmid and genomic DNA dilutions, we estimate that 1.3×10^6 copies / haploid genome of the SINE exist in the harbour seal genome (data not shown). Assuming a haploid 'C' value of 3×10^9 bp, SINEs constitute approximately 8.7% of the harbour seal genome. Estimates of SINE copy number based on the number of positive clones (12) isolated from the size-selected genomic library, however, are an order of magnitude lower than that determined by slot-blot hybridization analysis. Approximately 400 clones with inserts of 300−800 bp were screened representing $1.2-3.2\times10^5$ bp of the seal genome. Assuming that (i) the size-selected fraction cloned was a random, representative sample of the genome; (ii) SINEs are randomly distributed within this fraction; and (iii) they are evenly distributed throughout the genome, we estimate that a SINE sequence occurs every 10−27 kbp in the seal genome. Based on this analysis, approximately $1-3\times10^5$ copies of the SINE sequence are present in the seal genome. The markedly different estimates for the SINE copy number derived by the two methods employed may be related to the generalizing assumptions described above. SINE copy number in the harbour seal, therefore, either exceeds similar estimates in other mammalian orders (based on slot-blot analysis) such as primates (*Alu*—$3-5\times10^5$) and rodents (B1—1×10^5, B2—0.8×10^5) by a factor of 2−10 (5), or seal SINEs are of similar abundance in the genome (based on library screening) as other mammalian SINEs. If the relatively high copy number estimate of the harbour seal SINEs based on slot-blot analysis is correct, it may be explained in part by an association of SINEs with other repeated elements which underwent saltatory amplification, as has been suggested for satellite DNAs (24).

Evolutionary conservation of a SINE family in the Canoidea

The oligonucleotide derived from the canoid SINE sequences did not hybridize to genomic DNA from a fish or representatives of mammalian orders other than the Carnivora (data not shown). Hybridization of the oligonucletide to genomic DNA of canoid carnivores, but not to the cat (superfamily Feloidea), indicates that the SINE is restricted to the superfamily Canoidea (Figure 4). The Canoidea includes the families Otariidae and Ursidae in addition to the Mustelidae (mink), Canidae (dog, wolf, coyote) and Phocidae (seal). These results, and the sequence information available, contrast with the findings of Minnick *et al.* (20) who found no hybridization of a canid SINE to DNA from canoid representatives other than canids (i.e. mustelid [ferret] and ursid [bear]). Since the repeat is specific to the superfamily Canoidea, we suggest that the SINE be named the '*Can*' family repeat.

Using divergence times based on fossil evidence (25), the first *Can* SINE 'master copy' (10) probably arose at some time between the divergence of superfamilies Feloidea and Canoidea (approximately 55 million years ago [MYA]) and the divergence of the family Canidae from the other canoid families (50 MYA). The high copy number and relatively high sequence dissimilarity of *Can* family SINEs in the harbour seal support the hypothesis that it is a relatively ancient SINE. For example, the primate *Alu* family, which is also characterized by high copy number and sequence dissimilarity, arose approximately 60 MYA, while the more conserved B1, B2 and ID SINEs of the Rodentia are probably less than 40 million years old (11). Since the probe that detected SINEs in canoid species was designed to a conserved sequence block (average sequence similarity of 89%), the relative intensities of the seal, mink and canid hybridization signals, with seal and mink of greater intensity, suggest that the element is either more abundant or more highly conserved in the genomes of the closely related mustelids and phocids than it is in canids.

The pattern and rate of SINE retropostion may differ between families or even species. Kido *et al.* (12) have suggested that the pattern of amplification and dispersion of SINEs in closely related salmonids may have played a role in their speciation, perhaps by facillitating reproductive isolation. The proliferation of *Can* SINEs in canoidean genomes, as well as other SINE families in other mammalian orders, at the time of rapid mammalian ordinal and familial divergences 40−60 MYA, implies an evolutionary role for SINE retroposition in mammalian radiation, either as a cause or a consequence of speciation by directly acting on genome structure.

In summary, we have cloned SINE repeats from the harbour seal which are homologous to SINEs found in other carnivores (19−21). We propose that these elements constitute a family of SINEs specific to the superfamily Canoidea, the *Can* SINEs. While the element is predominately dispersed in the harbour seal genome, some elements may also be associated with a larger repeated element such as a satellite DNA. *Can* SINEs contain a CT repeat motif which may make them useful as a general source of polymorphic genetic markers similar to microsatellites for many applications.

ACKNOWLEDGEMENTS

We thank W.D.Bowen, S.Iverson, D.Boness and M.Muelbert for the collection of harbour seal blood samples. Mammalian DNA samples were also provided by H.Farid and K.Richard. Funding for this research was from a Natural Sciences and Engineering Research Council (NSERC) of Canada operating grant to JMW and a NSERC postgraduate scholarship to DWC. We thank an anonymous reviewer for comments, especially regarding copy number estimates of the seal SINE.

REFERENCES

1. Franck, P. C. , Harris, A. S., Bentzen, P., Denovan-Wright, E. M., and Wright, J. M. (1991) In Maclean, N. (ed.) *Oxford Surveys on Eukaryotic Genes*. Oxford University Press, Oxford, Vol. **7**, pp. 51–82.
2. Jelinek, W. R. and Schmidt, C. W. (1982) *Ann. Rev. Biochem.*, **51**, 813–844.
3. Wright, J. M. (1993) In Hochachka, P. W. and Mommsen, T. (eds.), *Biochemistry and Molecular Biology of Fishes*. Elsevier Press, Amsterdam, Vol. 2, pp 57–91.
4. Wright, J. M. (1994) *Genome*, **37**, 345–347.
5. Rogers, J. H. (1985) *Intern. Rev. Cytol.*, **93**, 187–279.
6. Deininger, P. L. (1989) In Berg, D. E. and Howe, M. M. (eds.) *Mobile DNA*. American Soc. Microbiol., Washington, pp. 619–636.
7. Okada, N. (1991) *Trends in Ecol. Evol.*, **6**, 358–361.
8. Singer, M. F. and Skowronski, J. (1985) *Trends Biochem. Sci.*, **10**, 119–122.
9. Hutchison, C. A., Hardies, S. C., Loeb, D. D., Shehee, W. R., and Edgell, M. H. (1989) In Berg, D. E. and Howe, M. M. (eds.), *Mobile DNA*. American Soc. Microbiol., Washington, pp. 593–617.
10. Deininger, P. L., Batzer, M. A., Hutchison, C. A., and Edgell, M. H. (1992) *Trends in Genet.*, **8**, 307–311.
11. Deininger, P. L., and Daniels, G. R. (1986) *Trends in Genet.*, **2**, 76–80.
12. Kido, Y., Aono, M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M., and Okada, N. (1991) *Proc. Nat. Acad. Sci. USA*, **88**, 2326–2330.
13. Yoshioka, Y., Matsumoto, M., Kojima, S., Ohshima, K., Okada, N., and Machida, Y. (1993) *Proc. Nat. Acad. Sci. USA*, **90**, 6562–6566.
14. Okada, N., K. Ohshima, K. (1993) *J. Mol. Evol.*, **37**, 167–170.
15. Ohshima, K., Koishi, R., Matsuo, M., and Okada, N. (1993) *Proc. Nat. Acad. Sci. USA*, **90**, 6260–6264.
16. Sambrook, J., Maniatis, T., and Fritsch, E. F. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
17. Goode, B. L., and Feinstein, C. (1992) *Biotechniques*, **12**, 374–375.
18. Altshul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.*, **215**, 403–410.
19. Lavrentieva, M. V., Rivkin, M. I., Shilov, A. G., Kobetz, M. L., Rogozin, I. B., and Serov, O. L. (1989) *Dokl. Akad. Nauk. SSSR*, **307**, 226–230.
20. Minnick, M. F., Stillwell, L. C., Heineman, J. M., and Steigler, G. L. (1992) *Gene*, **110**, 235–238.
21. Burnett, R. C. (unpublished) [Z25418]
22. Russo, T., Duilio, A., Ammeodola, R., Constanzo, F., and Cimino, F. (1987) *Nucleic Acids Res.*, **15**, 8562.
23. Gauss, D. H., Gruter, F., and Sprinzl, M. (1979) In Schimmel, P. R., Soll, D., and Abelson, J. N. (eds.) *Transfer RNA: Structure, Properties, and Recognition*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 520–538.
24. Brutlag, D. L. (1980) *Ann. Rev. Genet.*, **14**, 121–144.
25. Wayne, R. K., Benveniste, R. E., Janczewski, D. N., and O'Brien, S. J. (1989) In Gittleman, J. L. (ed.) *Carnivore Behaviour, Ecology, and Evolution*. Cornell Universiy Press, Ithaca, New York, pp. 465–494.