Genes Genet. Syst. (1999) **74**, p. 257–266

Sequence variation and evolution of the mitochondrial DNA control region in the musk shrew, *Suncus murinus*

Takahiro Yamagata*, and Takao Namikawa

Laboratory of Animal Genetics, Graduate School of Bioagricultural Sciences, Nagoya University, Furocho, Chikusa, Nagoya 464–8601, Japan

(Received 9 July 1999, accepted 13 November 1999)

The complete mitochondrial DNA (mtDNA) control region was cloned and sequenced in the musk shrew, *Suncus murinus*, Insectivora. The general aspect was similar to that found in other mammals. We have found in two locations of this region the presence of arrays of tandem repeats like those in other shrew species. One array was located in the left domain containing the termination-associated sequences (TAS) and the length of a copy was 77 bp. The other repeats were situated upstream from the recognition site for the end of H-strand replication in the right domain and were 20 bp long. The left halves of the control region containing the former repeats were sequenced and compared in several laboratory lines and wild animals from different localities, variations in copy number of repeated sequences were found both among individuals and within an individual. A comparative study of repeated sequences provides useful indication for the origin and evolution of tandem repeated sequences. Strand slippage and mispairing during replication of mtDNA with concerted manner is currently regarded as a dominant theory to account molecular mechanism for tandemly repeated sequences, and the pattern of sequence and length variation in our study supports this theory. Our results, however, suggest that the evolution of the repeated sequences containing the TAS in the musk shrew might go through the process of two steps; at the first step one complete repeated and several incomplete repeated sequences had reproduced in common ancestor of the shrew, and the second stage step-up of complete repeated sequences occurred with concerted evolution after differentiation into continental and insular groups.

INTRODUCTION

The control region is the only major noncoding region of mammalian mitochondrial DNA (mtDNA) and the sequence variation of this segment is higher than those of other regions. Thus, this region has been sequenced and compared in many animal species to consider inter- or intraspecies relationships. Sequences of the control region determined in many species indicate that its structure is almost common in mammals: this region can be divided into three domains and the central region is relatively conserved. On the other hand, the left domain adjacent to the Pro-tRNA gene, containing the termination-associated sequences (TASs or ETASs: MacKay et al., 1986; Madsen et al., 1993b; Sbisà et al., 1997), and the right domain adjacent to the Phe-tRNA gene that contains the site of initiation for H-strand replication (O_H) and the conserved sequence blocks (CSBs) have evolved rapidly (Brown et al., 1986; Saccone et al., 1991). The variation in number of tandem repeats which exist in the control region have reported to contribute to extensive size variation and heteroplasmy of mtDNA in many animal species (e.g., Buroker et al., 1990; Johansen et al., 1990; Biju-Duval et al., 1991; Hayasaka et al., 1991; Wilkinson and Chapman, 1991; Stewart and Baker, 1994; Fumagalli et al., 1996; Prager et al., 1996; Wilkinson et al., 1997; Douzery et al., 1997).

The musk shrew (*Suncus murinus*) is a small mammalian species belonging to order Insectivora and has been domesticated as an experimental animal. As musk shrews are widely distributed in East, Southeast and South Asia, Southwest islands in the Pacific Ocean and East Africa, we have caught them from some of these areas, maintained as local populations, and established several laboratory lines. We have examined genetic relationship among local shrew populations by the restric-

* Corresponding author.

tion fragment length polymorphism (RFLP) analysis of mtDNA and evaluated that musk shrews are differentiated with intersubspecific level between from the continent in South Asia and from islands in Southeast Asia (Yamagata et al., 1990). In those studies size variation was found in homologous restriction fragments among individuals and any of these fragments were suggested to contain the control region. We determined the sequences of the control region and compared among lines and individuals originated from different localities in the present study. Fumagalli et al. (1996) reported the presence of two kind of tandem repeats in the control region in two genera of shrews, *Sorex* and *Crocidura*, and in our research similar repeated sequences were found in *Suncus murinus* classified into same Soricidae family. We discussed the origin and mechanism for replication of tandem repeats commonly found in insectivore animals.

MATERIALS AND METHODS

Specimens examined. Specimens were collected from 7 individuals of the 6 different musk shrew lines maintained at our laboratory (NAG, TKU and OKI from Japan, BAN from Bangladesh, SRI from Sri Lanka and KAT from Nepal) and two wild animals from India and Pakistan (Table 1). Seven samples from the lines have been analyzed by RFLP and classified into 6 mtDNA haplotypes (Yamagata et al., 1995).

DNA preparation, cloning, polymerase chain reaction (PCR) and DNA sequencing. The mtDNA sample of an individual from each of shrew line was extracted by using the alkaline lysis procedure as described by Tamura and Aotsuka (1988). From two wild individuals of the shrew, total DNA was isolated from liver fixed in 95% ethanol. Ethanol-fixed tissue was rinsed and homogenized in 5% Citric acid buffer. The homogenate was centrifuged and total DNA was prepared by conventional methods.

BAN-108A mtDNA sample, sequenced complete control region, was digested by *Cla*I. The fragments were ligated with plasmid vector pUC119 cut by *Acc*I and transformed into competent MV1184. The plasmid inserted with the fragment containing control region was extracted and selected. For sequencing, different sized DNAs were subcloned by using the Kilo-Sequencing Deletion kit (Takara Shuzo Co.) from this selected clone. By inoculating the helper phage M13KO7 into the MV1184 cells with these subcloned vectors, single-stranded DNAs were produced for DNA sequencing templates. After the concentration of template was adjusted, 6 μl of the template were sequenced using the Dye Primer Cycle Sequencing kit (Perkin-Elmer Co.) with –21M13 universal dye primers. The product was run and the sequence was read automatically using the DNA Sequencer, Model 373A (Perkin-Elmer Co.).

The above first analysis allowed us to characterized the locations and structures of two kind of repeats (LR and SR, Fig. 1) in the control region of the musk shrew. In

Table 1. List of specimens of the musk shrew examined

Line	Individual No.	Locality (Captured year)
NAG	50B	Nagasaki, Japan (1973)
TKU	M205	Tokunoshima, Japan (1982)
OKI	CCCXLVIII	Okinawa, Japan (1983)
BAN	108A, 220B	Mymensingh, Bangladesh (1983)
SRI	I	Koralawella, Sri Lanka (1984)
KAT	89c	Katmandu, Nepal (1989)
(wild)	Pak-176	Islamabad, Pakistan (1991)
(wild)	Ind-308	Bhubaneswal, India (1992)

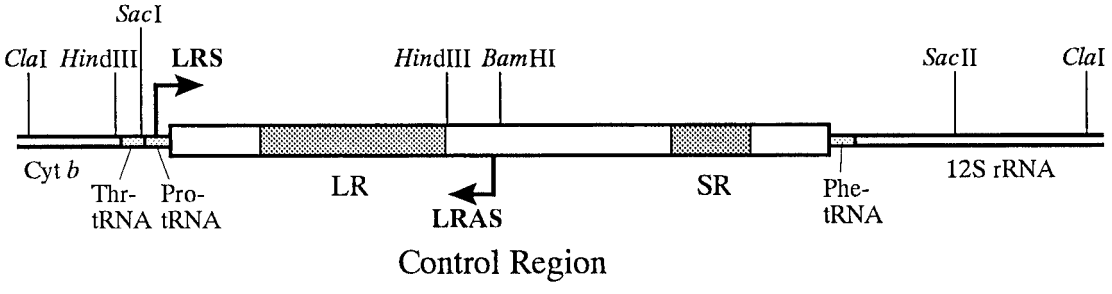

Fig. 1. Schematic diagram of the control region in the musk shrew mtDNA. Two dotted boxes in the control region indicate the long repeat (LR) and short repeat (SR). Vertical lines and arrows identify the restriction sites and the location of primers (LRS and LRAS) for PCR amplification of the region containing LR, respectively.

order to compare the repeats, a part of control region containing the long repeat (LR) was amplified by PCR and sequenced in the other mtDNA and total DNA samples. The primers used were LRS (5' CCCACCATCAACACCCAAAGC 3') and LRAS (5' AAGGGTTGCTGGTTTCACG 3'). The temperature profile for 25 cycles of amplification was 30 sec at 94°C, 30 sec at 62°C and 1 min at 72°C. After the amplification, the size of product was estimated by 2% agarose gel electrophoresis. The product was ligated with pGEM®-T vector (Promega Co.) and transformed into MV1184. Destined cloning vectors were selected and single-stranded DNAs were prepared for sequencing templates by inoculation of helper phage. The sequencing method was described as the above. More than 3 clones were sequenced for each shrew sample and the sequence determined as same in more than 2 clones was recognized true sequence of each sample.

DNA sequence analysis. DNA sequences were analyzed using the Phylogeny Inference Package (PHYLIP, Felsenstein, 1995). For estimate genetic distance among the control regions of musk shrews, sequences commonly determined (regions amplified by PCR except the LR array, 367–422 bp) were aligned. Then the genetic distance among specimens was calculated by the Kimura-2-parameter method (Kimura, 1980) using the Dnadist program in PHYLIP. Sequence data of *C. russula* (Fumagalli et al., 1996) was used as an outgroup. The phylogenic tree among shrews was constructed by the neighbor-joining method (Saitou and Nei, 1987) using the program in PHYLIP.

We also calculated the distance among all pairs of repeats within genomes and that among all pairs of repeats located at same positions among shrews. Potential secondary structures of tandem repeats were determined by the method of Zuker and Stiegler (1981) and illustrated using the genetic analysis program of the GENETYX (Software Develop. Co.). In order to assess the evolutionary dynamics of tandem repeats, we aligned all of tandem repeats irrespective of within and among genomes and conducted a maximum-persimony analysis in PHYLIP using two repeats of *C. russula* as an outgroup. To reduce the magnitude of the data set all identical repeats were classified as a single unit.

RESULTS

Control region of the musk shrew. The length of complete control region in BAN-108A mtDNA sample was 1,751 base pairs and the sequence is aligned in Fig. 2. General aspect of the control region in the musk shrew was similar to that in other vertebrate reported. The termination-associated sequences (TASs) and the conserved sequence blocks (CSBs) which are apparently important to the regulation processes of mtDNA replication, were determined by comparing the sequence in the shrew with other species sequences (Fumagalli et al, 1996; Wilkinson et al, 1997). The repetitive sequences were found in two potential locations of the control region in the musk shrew. The first one containing the TAS was located in the left domain and the second one was located between the conserved sequence blocks CSB1 and CSB2.

For the rest of the samples, a left half of control region containing the long repeat (LR) was sequenced. The electrophoresis after the amplification of samples by PCR showed two and three different sized bands in the SRI-I mtDNA and Pak-176 total DNA samples, respectively. Sequence analyses indicated these length differences were due to different copy numbers of the LR, i. e., two different fragments of the SRI-I sample had each of 6 and 7 copies of the repeats and three fragments of the Pak-176 possessed each of 6, 7 and 8 copies of the repeats. The RFLP analysis of the SRI-I mtDNA supported that this sample had two kinds of mtDNA of different sizes in the control region. These two samples of the shrews were expected to be heteroplasmic for two and three different sized DNA molecules. Other samples examined were homoplasmic.

Relationships among musk shrews from different localities. The phylogenic tree among musk shrews from different localities was constructed by the neighbor-joining method based on the sequences of the control region commonly determined (367–422 bp) using the data of *C. russula* as an outgroup and shown in Fig. 3. Nomenclatures in parentheses in Fig. 3 represent haplotypes of shrew mtDNA designated by RFLP analysis (Yamagata et al., 1995), OKI and TKU samples were identified to same SEA-C type and two kind haplotypes were found in BAN line by our previous study. The result from sequences of the control region was similar with the result by RFLP analysis; mtDNA haplotypes were divided into the islands' and continental groups; Japanese lines (NAG, OKI and TKU) and SRI belong to the former group and BAN and KAT belong to the latter group. The musk shrews from Pakistan and India have been expected to belong the continental group and their phylogenetic positions were consistent with localities of samples. However the relationship among localities within each group did not become so manifest by our present data except that OKI and TKU were very close both geographically and phylogenetically (all of animals from these two localities were also possessed same haplotype by the RFLP analysis, Yamagata et al., 1990).

Repeated sequences in the left (LR) and right (SR) domains. All individuals that the left domain of the control region was sequenced showed to possess long repetitive sequences (LR) containing the TAS. It consisted of five copies of a 77 bp sequence repeated in tandem, a 74

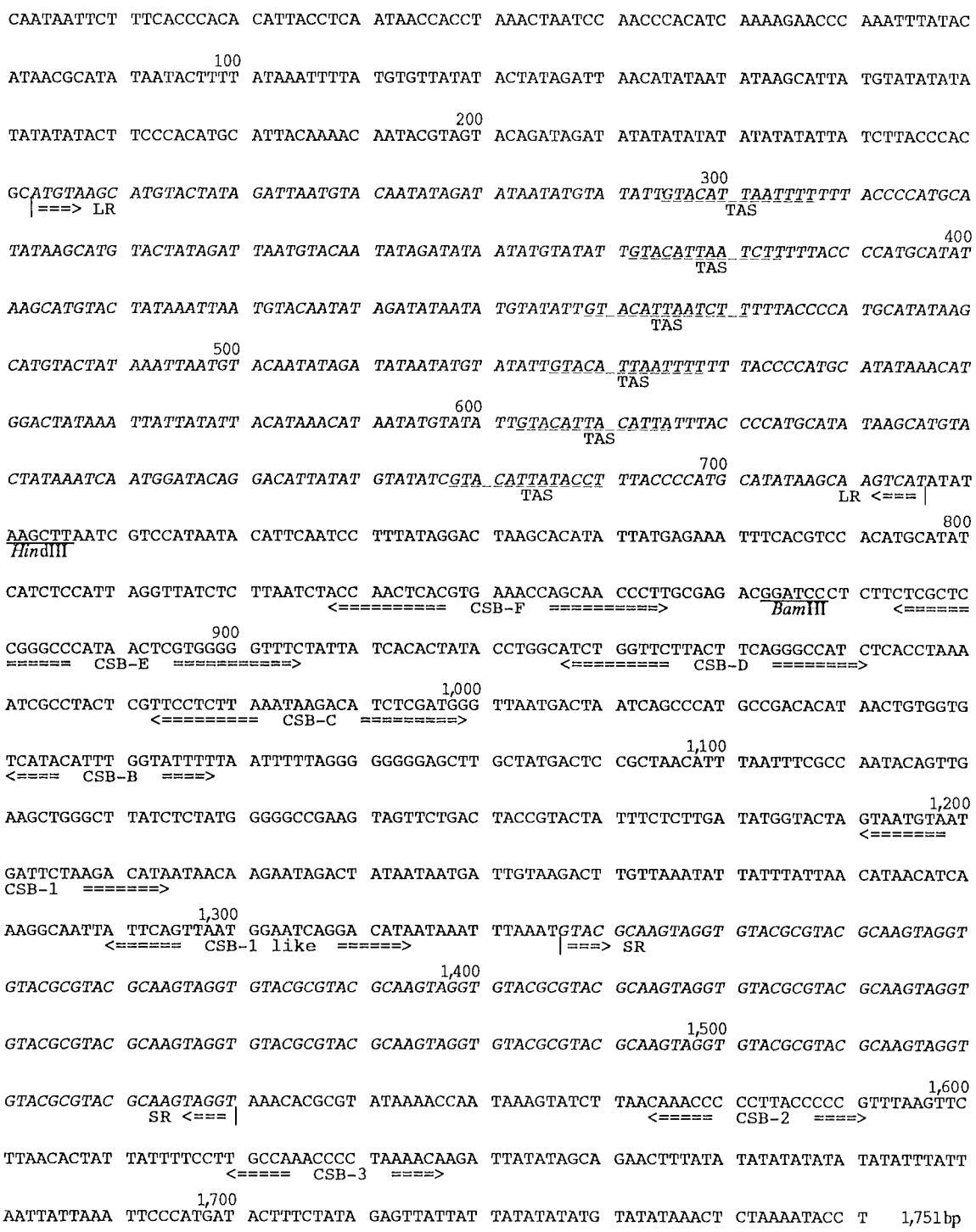

Fig. 2. Sequence of the mtDNA control region from *Suncus murinus*. The sequence is given for BAN-108A specimen and presented from 5' to 3' on the light strand. The long (LR) and short (SR) repeated sequences are indicated in italics. Sequences corresponding to the conserved sequence blocks (CSB B–F, 1–3) are delimited by '<= =>', and termination associated sequences (TASs) are also shown with broken underlines.

bp sequence and a largely deleted 15 bp sequence in BAN-108A sample that was sequenced in the complete control region (Fig. 2). The 3' flanking 77 bp repeat in 5 copies of 77 bp repeats had pretty different base composition from the rest copies of 77 bp repeats. So this 77 bp repeat and different sized repeats (74 bp and 15 bp) were regarded as 'incomplete' repeats and the rest of 4 copies of 77 bp repeats were regarded as 'complete' repeats. That is,

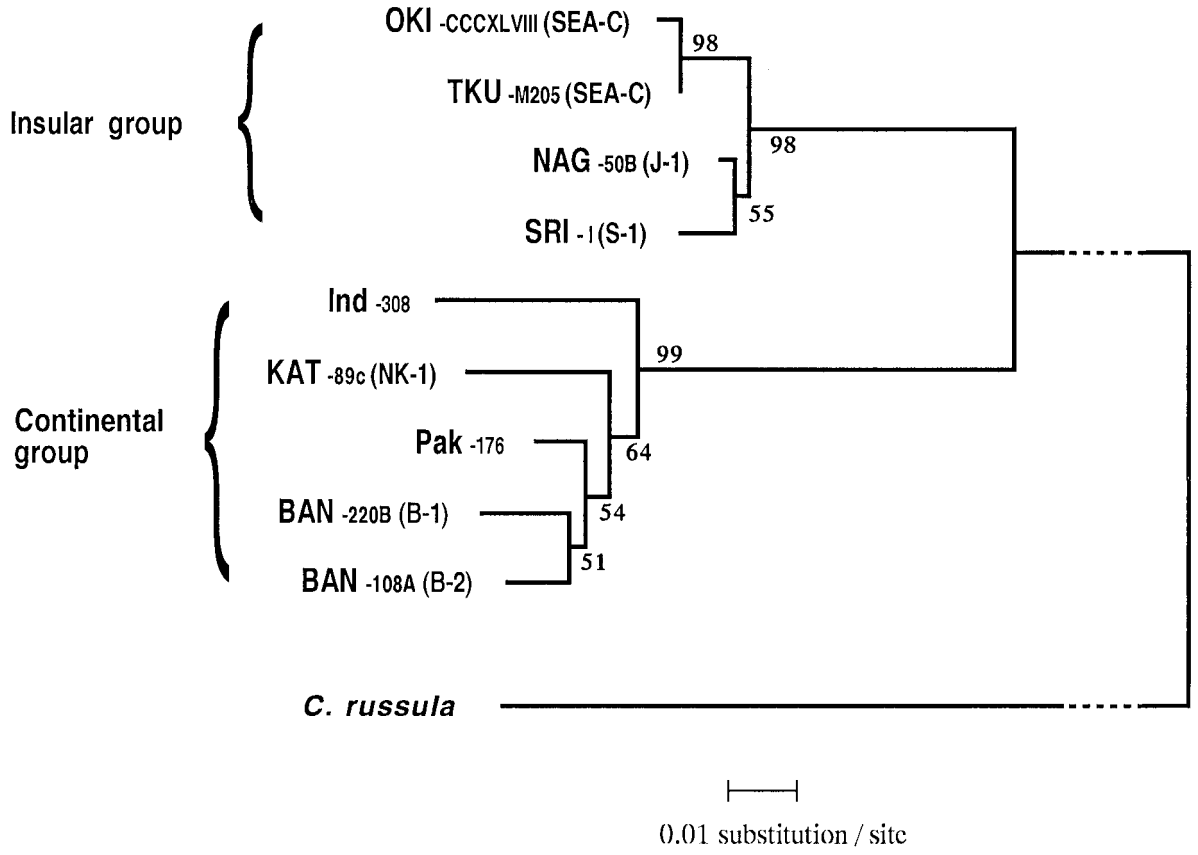

Fig. 3. A phylogenetic tree among musk shrews deduced from sequences of the mtDNA control region commonly determined (367–422 bp) by neighbor-joining method. Numbers indicate percentage bootstrap figures from 1,000 replications of the data. Nomenclatures in parentheses after specimen names and two groups indicated on the left of the figure represent haplotypes of shrew mtDNA designated by RFLP analysis and their classified groups, respectively (Yamagata et al., 1995). The data of *C. russula* as an outgroup was obtained from Fumagalli et al. (1996).

BAN-108A had in order from 5' end largely deleted 15 bp sequence, 4 copies of 77 bp complete repeats, 77 bp and 74 bp incomplete repeats. In other samples the basic organization of the LR array was same as that of BAN-108A sample. However, copy numbers of complete repeats and sizes of incomplete repeats were variable among individuals. The 5' flanking incomplete repeats were 15 bp sequences extensively deleted at the 5' end in all continental shrews (BAN, KAT, Ind and Pak), but those of insular shrews (NAG, OKI, TKU and SRI) were partially deleted 70–72 bp sequences. The 3' flanking incomplete repeats were 74–75 bp and 76 bp sequences in continental and insular shrews, respectively. The copy number of complete repeats was basically 4 in all individuals examined whereas variations of 3–5 copies were found in heteroplasmic samples, SRI-I and Pak-176. These alignments are shown in Fig. 4 assigning a number in order from the 5' end (the 5' flanking incomplete repeat was numbered 1'st for convenience' sake).

The short repeated sequences (SR) only examined in BAN-108A sample were located between the conserved sequence blocks CSB1 and CSB2. The length and number of the repeated motif were 20 bp and 10, and the sequence of a motif was '$(GTACGC)^2$ AAGTAGGT'. No sequence variations were found among the motifs (Fig. 2).

Relationships among repeats within and among genomes. Among individuals of musk shrews, nucleotide diversities between pairs of repeats situated on same numbered position were calculated (total copy number of the repeats was unified to 7 for homoplasmic samples). Averages of the distances among pairs between continental and insular shrews (*C* vs. *I*) and between continental or insular fellows (*C* vs. *C* / *I* vs. *I*) were shown in Fig. 5. There were significant differences in nucleotide diversities between *C* vs. *I* and *C* vs. *C* / *I* vs. *I*, but no significant differences among different located repeats in each pair of shrews. The range of average values for each distance was 0.029 to 0.045 substitution / site among *C* vs. *C* / *I* vs. *I* and that among *C* vs. *I* was 0.098 to 0.138 substitution / site.

We also calculated nucleotide diversities among all pairs of repeats within genomes and indicated them in

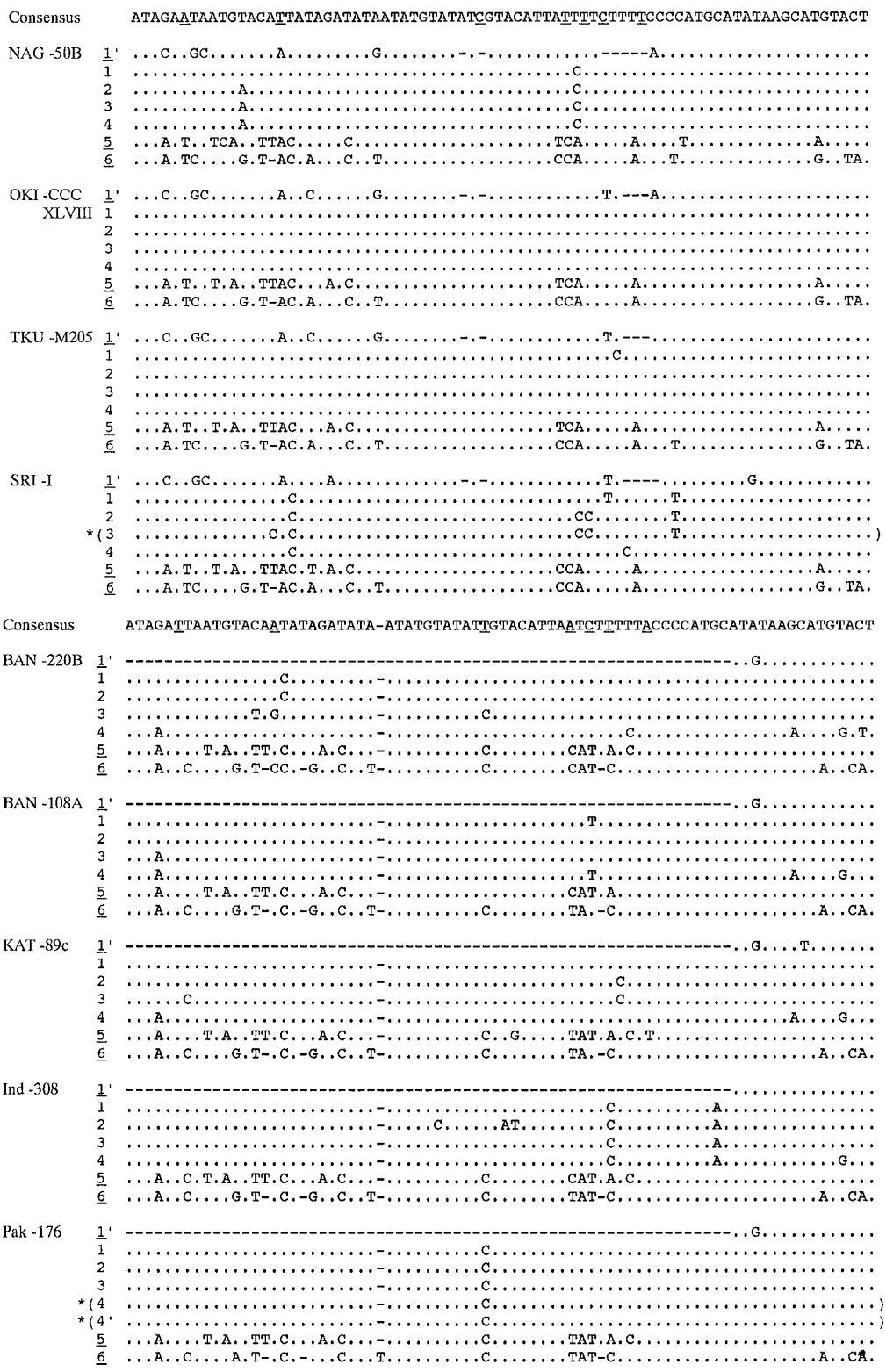

Fig. 4. Alignments of the LR tandem repeats from 9 specimens of the musk shrew. Nucleotides found most frequently among specimens are regarded as consensus and presented on the top of the alignment. Some consensus nucleotides were different between insular and continental shrew specimens and shown by underlined letters. Dots indicate nucleotide identity and dashes are proposed indels as consensus. The repeat numbers are assigned in order from the 5' end and indicated by normal and underlined letters for complete and incomplete repeats, respectively. Sequences within parentheses of SRI-I and Pak-176 (indicated by asterisks) are alternative for heteroplasmy; two kind of different sized molecules of SRI-I mtDNA have the third repeat or not, and three kinds of Pak-176 mtDNA molecules have both the 4th and 4'th repeats, the only 4th repeat or neither repeats.

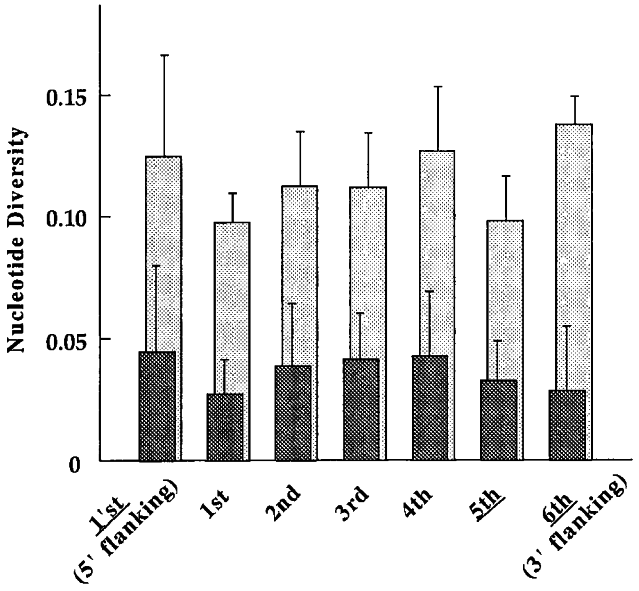

Fig. 5. Nucleotide diversities between pairs of repeats located same positions among the musk shrew. Gray bars and black bars indicate averages of distances among each pair between continental and insular shrews (*C* vs. *I*) and between continental or insular fellows (*C* vs. *C* / *I* vs. *I*), respectively. Vertical line on each bar indicates the standard deviation. See Fig. 4 for the position number.

Fig. 6. Within genomes there were extensively large distances between complete (1st, 2nd, 3rd and 4th) and incomplete (1'st, 5th and 6th) repeats, and among incomplete repeats, whereas the sequences were very similar to each other among complete repeats. The average values and the standard deviations for distances were 0.023 ± 0.027 and 0.202 ± 0.060 substitution / site among complete repeats, and between complete and incomplete repeats / among incomplete repeats, respectively. The former value was smaller than the value for distances among *C* vs. *C* / *I* vs. *I* shrews and the latter value was larger than that among *C* vs. *I*.

Secondary structure of repeats. Each single-strand molecule of complete repeats in the LR array was capable of forming the energetically stable secondary structure. The most favored secondary structures differ slightly among individuals (Fig. 7), and the free energy associated with the formation ranges approximately –19.5 to –32.9 kcal / mol. The repeats of the SR array could also form stable secondary structure. As described for other species (Stewart and Baker, 1994; Fumagalli et al., 1996), the free energies of their repeats folding second structures are increasing according to copy number of repeats: for example, one repeat of the SR array has folding energy of –4.6 kcal / mol, and the 2-, 3- and 4-repeat arrays have those of –16.5, –29.1 and –41.0 kcal / mol, respectively.

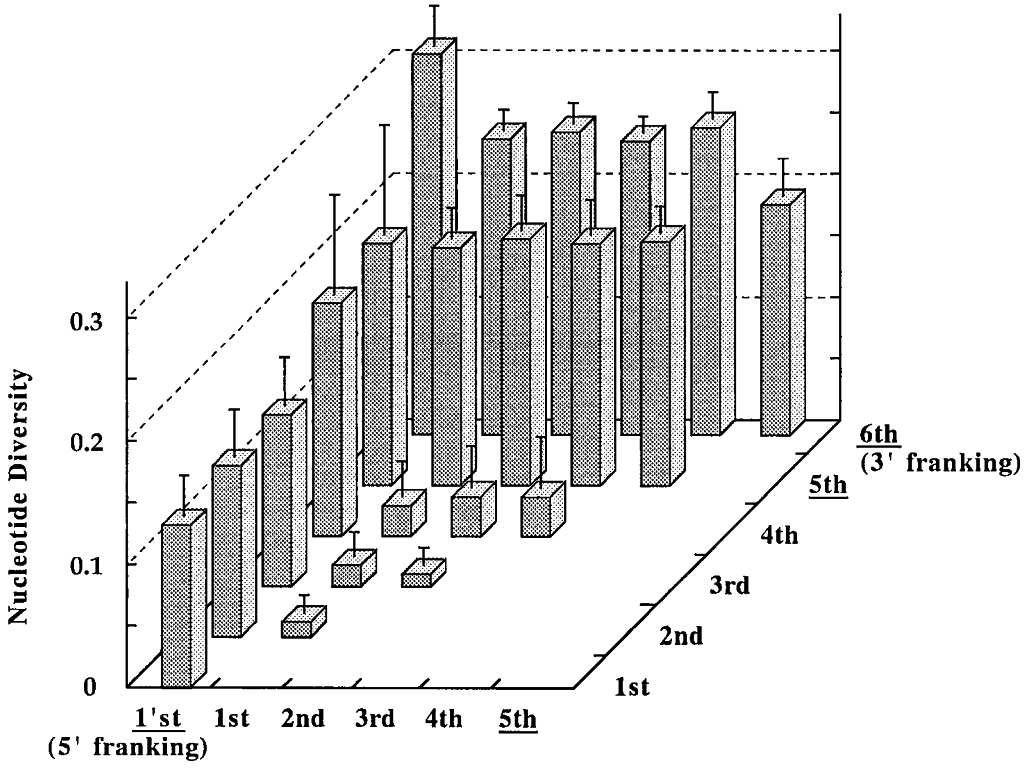

Fig. 6. Mean nucleotide diversities between pairs of each repeat within genomes calculated from 9 shrew specimens. Vertical line on each box indicates the standard deviation. See Fig. 4 for the position number.

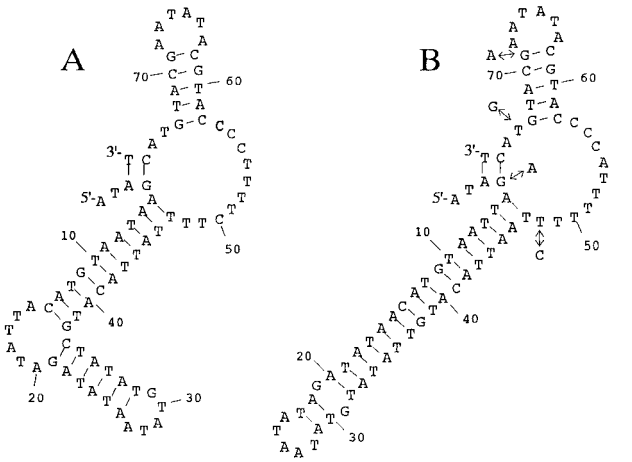

Fig. 7. Examples of favored secondary structures of the LR repeats. (A) Folded structure of one copy of the complete repeated sequence from OKI-CCCXLVIII (free energy = –25.1 kcal / mol). (B) Folded structure of that from BAN-108A (free energy = –30.8 kcal / mol). Double-headed arrows mark the polymorphic positions among the complete repeats.

DISCUSSION

The complete mtDNA control region was sequenced in an individual of the musk shrew. The sequences of conserved elements, which thought to play a major role in the regulation of the replication processes of mtDNA (TASs and CSBs), were observed in this region (Fig. 2). Moreover two kinds of tandemly repeated sequences were found at potential locations where repeats tend to occur – in the left domain containing TASs (LR array) and between CSB1 and CSB2 (SR array). The simultaneous existence of both these repeats was consistent with that observed in *Sorex* and *Crocidura* belonging to the same Soricidae family (Fumagalli et al., 1996). The comparative analysis of sequences of the left half of the control region indicated that the region from 5' end to the LR array was more variable although the downstream of the LR array was conservative. These results suggest that the general organization of the mtDNA control region of the musk shrew is similar to that of other mammalian species.

The genetical relationship among lines or wild populations of the musk shrew was discussed from the RFLP analysis of mtDNA (Yamagata et al., 1995). The result of this study from sequences of the control region supported the idea predicted by the RFLP analysis that shrew populations were generally diverged into the insular and continental groups (Fig. 3), though each number of samples was only one from lines and populations. Two specimens surveyed from wild animals were prepared as total DNA samples and could not be used as materials for the RFLP analysis. It might be speculative to discuss phylogenetic relationship only from comparison of specific genes or elements, and as the sequenced region is a non-protein coding region which contain much insertion/deletion mutations it is difficult to assess relative ratio of base displacement to insertion/deletion mutations for phylogenetic analysis. Therefore these two animals from India and Pakistan at least belong to the continental group, but the relationship between them and other shrews in the continental group was not clear.

The heteroplasmy of mitochondrial genome has been reported for numerous animals and many of them were attributable to variation in number of tandem repeats in the control region (Arnason and Rand, 1992; Buroker et al., 1990; Fumagalli et al., 1996; Wilkinson and Chapman, 1991). Musk shrews were also found to have variant sized molecules with different copy numbers of tandem repeats within an individual and the frequency of heteroplasmy was 22.2%. Two individuals indicated heteroplasmy (SRI-I and Pak-176) were found to have two and three different molecules, respectively, by PCR analysis. Fumagalli et al. (1996) reported, however, all different sized bands found by Southern blot analysis were not detected by PCR analysis, and that is presumably due to the much higher sensitivity to detection of transfer-hybridization methods compared with ethidium bromide-stained agarose gels and to selective amplification of particular repeat sizes with PCR. So musk shrews indicated heteroplasmy are possible to have more different molecules besides detected bands in this study. Moreover among variant sized molecules within an individual, some other differences of sequences were observed in addition to the different number of tandem repeats (variation of copy numbers in two or three TA repeat arrays found at the 5' end from the LR array, data was not shown). More detailed sequence analysis within an individual is necessary to clear relationship between different molecules concerning to heteroplasmy.

The general composition of the LR array was analogous to tandem arrays located same positions reported in other vertebrate species such as the sturgeon (Buroker et al., 1990), the cod (Johansen et al., 1990; Arnason and Rand, 1992), the tree frog (Yang et al., 1994), the evening bat (Wilkinson and Chapman, 1991; Wilkinson et al, 1997) and several species of the shrew (Stewart and Baker, 1994; Fumagalli et al., 1996), basically consisting of one to several copies of tandem repeats and single incomplete repeat which were related but distinct from tandem repeats adjacent at the 3' end of the array. The musk shrew was, however, found to possess two copies of incomplete repeats at the 3' end of the array and one copy at the 5' end (large deletion was observed in the continental shrews), and the genetic distances among these incomplete repeats were similar to those between complete and incomplete repeats (Fig. 6). Buroker et al. (1990) presented repeat number will increase during the replication of mtDNA, when a repeat in the D-loop strand forms a secondary structure, competes with and displaces the H-strand for pairing with the L-strand, and is extended

into the nascent H-strand by replication. And they predicted concerning their model that there must be at least three repeats, two repeats at both end of the array could diverge and the central repeats evolve in a concerted manner. The musk shrew surveyed in this work satisfied these predicted conditions such as a capability of forming secondary structure of the repeat, possessing over three distinct repeats and etc., supporting their model.

Several molecular mechanisms have been proposed to account for increase and decrease of tandem repeats, and currently favored proposed mechanism is strand slippage and mispairing during replication (Buroker et al., 1990; Hayasaka et al., 1991; Wilkinson and Chapman, 1991; Madsen et al., 1993a). In the musk shrew variation of repeat number was also assumed to be caused by unidirectional replication slippage (Fumagalli et al., 1996). However considering the comparison of the LR tandem array among individuals, we estimate there were several steps of events on the productive process of the LR array at different periods. The sequence differences were considerably small among complete tandem repeats that evolve in a concerted manner, but the genetic distances among three incomplete repeats and between complete and incomplete repeats within an individual were larger than those between homologous repeats of insular and continental shrews that were differentiated with an intersubspecific level (Fig. 5 and 6). In order to verify these we aligned all tandem repeats in the LR array and analyzed them by a maximum-persimony analysis (as incomplete repeats at the 5' end of the continental shrews were extremely deleted, they were excluded from this analysis). As a result two kind of incomplete repeats at the 3' end (5th and 6th) and complete repeat (1st–4th) were proved to have diverged before insular and continental musk shrews were differentiated, but relationship among complete repeats in each of two types of shrews did not reveal clear clades (Fig. 8). Replication of complete repeats was only suggested to occur after musk shrews have been diverged to insular and continental groups. Therefore we build up the following hypothesis for a production process of tandem repeats in the musk shrew: existing incomplete repeats had been duplicated (to at least 4 copies) at the first step in common ancestor, and after divergence to insular and continental groups tandem repeats have subsequently increased in copy number at the second step with concerted manner of evolution. Although the following was not described by Fumagalli et al. (1996), sequences assumed to be a possible part of repeats were observed adjacent at both 5' and 3' ends of the repeat array (R1 array in their paper) in *Crocidura*, but were not found in *Sorex*. So the duplication of three incomplete repeats at both ends of the array might have occurred commonly in *Suncus* and *Crocidura* belonging to Crocidurinae subfamily. However the distances between complete and incomplete repeats in each genus were smaller than those between homologous repeats in *Suncus* and *Crocidura* (Fig. 8). The first step of duplication of tandem repeats was not thought to be occurred in a common ancestor of *Suncus* and *Crocidura*, but the concerted evolution of repeats was rather caused after differentiation of the genus. Another survey of variation within species was necessary in other animals in Crocidurinae subfamily, but at least this construction of the tandemly repeated array was unique for these shrew species. Furthermore stability of secondary structures of complete repeats was slightly higher than that of incomplete repeats (averages of free energies were –26.7, –16.2 and –19.4 kcal / mol in the 1st–4th, 5th and 6th repeats of the array, respectively), suggesting that the two-step model for the generation of repeated sequences was applicable to the evolution of repeats in *Suncus*.

The copy numbers of tandem repeats in the LR array

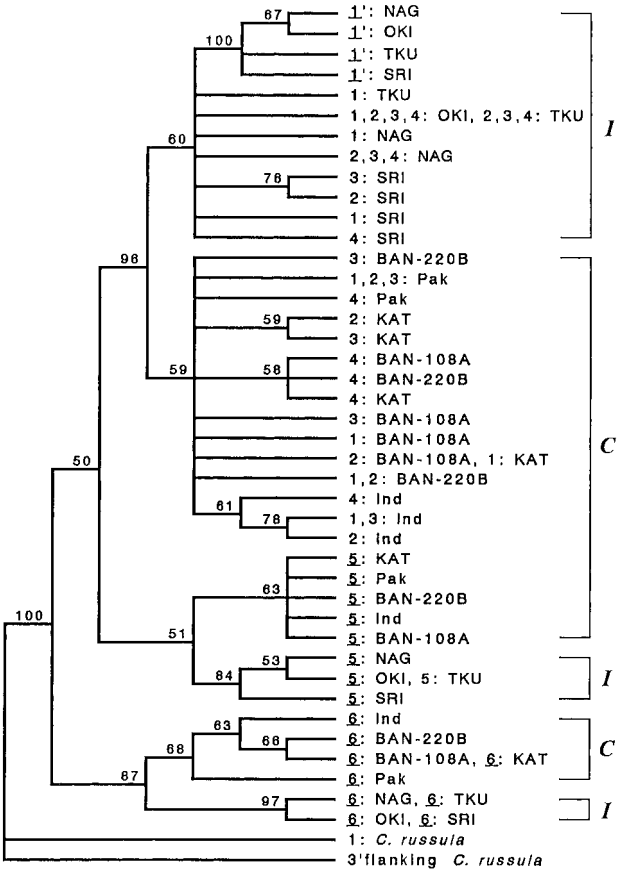

Fig. 8. Phylogenetic relationships among tandem repeats of the LR in the control region by maximum-persimony analysis. All identical repeats were classified as a single unit. Information in each repeat is indicated as repeat number (See Fig. 4) and specimen name (specimen number is omitted except for the BAN samples). On the right of the figure it shows that each repeat belongs to whitch type of shrew, continental (*C*) or insular (*I*). The data of *C. russula* (Fumagalli et al., 1996) was used as an outgroup. The tree presented is the 50%-majority-rule consensus of 100 bootstrap trees and numbers indicate percentage bootstrap figures.

without incomplete repeats were four in all individuals except heteroplasmic animals possessed 3 to 5 copies. Because differences were larger among homologous repeats between individuals than among tandem repeats within an individual, a high rate of concerted evolution was suggested within genomes in each individual or population. Nevertheless the basic copy numbers were same in all animals; it is interesting to discuss this point. As the number of surveyed animals was only nine, this is apparently an unexpected result, but it was anticipated the possibility of some action forced to make up the number of copies or the lengths of the repeat array. It is desirable to investigate the sequences of tandem repeats in more specimens containing also the SR array.

We are indebted Dr. H. Ikeda, National Institute of Animal Health and Dr. K. Tsuchiya, Miyazaki Medical College for the collection and DNA preparation of Indian and Pakistan shrew samples. The nucleotide sequence data reported in this paper will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession numbers AB011390 - AB011398.

REFERENCES

Arnason, E. and Rand, D. M. (1992) Heteroplasmy of short tandem repeats in mitochondrial DNA of Atlantic cod, *Gadus morhua*. Genetics **132**, 211–220.

Biju-Duval, C., Ennaffa, H., Dennebouy, N., Monnert, M., Mignotte, F., Soriguer, R. C., El Gaaïed, A., El Hili, A. and Mounolou, J. -C. (1991) Mitochondrial DNA evolution in lagomorphs: origin of systematic heteroplasmy and organization of diversity in European rabbits. J. Mol. Evol. **33**, 92–102.

Brown, G. G., Gadaleta, G., Pepe, G., Saccone, C. and Sbisà, E. (1986) Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. J. Mol. Biol. **192**, 503–511.

Buroker, N. E., Brown, J. R., Gilbert, T. A., O'Hara, P. J., Beckenbach, A. T., Thomas, T. K. and Smith, M. J. (1990) Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics **124**, 157–163.

Doda, J. N., Wright, C. T. and Clayton, D. A. (1981) Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA **78**, 6116–6120.

Douzery, E. and Randi, E. (1997) The mitochondrial control region of Cervidae: Evolutionary patterns and phylogenetic content. Mol. Biol. Evol. **14**, 1154–1166.

Felsenstein, J. (1995) *PHYLIP* (Phylogeny Inference Package) ver. 3.572. Univ. of Washington, Seattle.

Fumagalli, L., Taberlet, P., Favre, L. and Hausser, J. (1996) Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. Mol. Biol. Evol. **13**, 31–46.

Hayasaka, K., Ishida, T. and Horai, S. (1991) Heteroplasmy and polymorphism on Japanese monkey: association with tandemly repeated sequences. Mol. Biol. Evol. **8**, 399–415.

Johansen, S., Guddal, P. H., and Johansen, T. (1990) Organization of the mitochondrial genome of Atlantic cod, *Gadus morhua*. Nucleic Acids Res. **18**, 411–419.

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**, 111–120.

MacKay, S. L. D., Olivo, P. D., Laipis, P. J., and Hauswirth, W. W. (1986) Template-directed arrest of mammalian mitochondrial DNA synthesis. Mol. Cell. Biol. **6**, 1261–1267.

Madsen, C. S., Ghivizzani, S. C., and Hauswirth, W. M. (1993a) *In vivo* and *in vitro* evidence for slipped mispairing in mammalian mitochondria. Proc. Natl. Acad. Sci. USA **90**, 7671–7675.

Madsen, C. S., Ghivizzani, S. C., and Hauswirth, W. M. (1993b) Protein binding to a single termination-associated sequence in the mitochondrial DNA D-loop region. Mol. Cell. Biol. **13**, 2162–2172.

Prager, E. M., Tichy, H., and Sage, R. D. (1996) Mitochondrial DNA sequence variation in the Eastern house mouse, *Mus musculus*: comparison with other house mice and report of a 75-bp tandem repeat. Genetics **143**, 427–446.

Saccone, C., Pesole, G., and Sbisà, E. (1991) The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. J. Mol. Evol. **33**, 83–91.

Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**, 406–425.

Sbisà, E., Tanzariello, F., Reyes, A., Pesole, G., and Saccone, C. (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene **205**, 125–140.

Stewart, D. T., and Baker, A. J. (1994) Patterns of sequence variation in mitochondrial D-loop region of shrews. Mol. Biol. Evol. **11**, 9–21.

Tamura, K., and Aotsuka, T. (1988) Rapid isolation method of animal mitochondrial DNA. Biochem. Genet. **26**, 815–819.

Wilkinson, G. S., and Chapman, A. M. (1991) Length and sequence variation in evening bat D-loop mtDNA. Genetics **128**, 607–617.

Wilkinson, G. S., Mayer, F., Kerth, G., and Petri, B. (1997) Evolution of repeated sequence arrays in the D-loop region of bat mitochondrial DNA. Genetics **146**, 1035–1048.

Yamagata, T., Ohishi, K., Faruque, M. O., Masangkay, J. S., Ba-Loc, C., Vu-Binh, D., Mansjoer, S. S., Ikeda, H., and Namikawa, T. (1995) Genetic variation and geographic distribution on the mitochondrial DNA in local populations of the musk shrew, *Suncus murinus*. Jpn. J. Genet. **70**, 321–337.

Yamagata, T., Tanaka, Y., Ishikawa, A., Namikawa, T., and Tomita, T. (1990) Genetic relationship among the musk shrews, *Suncus murinus* Insectivora, inhabiting islands and the continent based on mitochondrial DNA types. Biochem. Genet. **28**, 185–195.

Yang, Y. -G., Lin, Y. -S., Wu, J. -L., and Hui, C. -F. (1994) Variation in mitochondrial DNA and population structure of the Taipei treefrog *Rhacophorus taipeianus* in Taiwan. Mol. Ecol. **3**, 219–228.

Zuker, M., and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acid Res. **9**, 133–148.