

Letter to the Editor

Recombination in Animal Mitochondrial DNA: Evidence from Published Sequences

Emmanuel D. Ladoukakis and Eleftherios Zouros

Department of Biology, University of Crete, Greece; and Institute of Marine Biology of Crete, Iraklio Crete, Greece

The question of whether animal mitochondrial DNA (mtDNA) undergoes recombination has recently been the subject of intense debate. In a recent paper (Ladoukakis and Zouros 2001), we presented evidence for mtDNA recombination from direct recovery of recombinant molecules in the gonads of several males of the mussel *Mytilus galloprovincialis*. Because in this species there exist two lineages of mitochondrial genomes, each with a different mode of transmission (a phenomenon known as doubly uniparental inheritance [DUI] of mtDNA; Zouros et al. 1994), the possibility existed that mtDNA recombination was a peculiarity of DUI. Here, we present evidence for recombination in three different species with standard maternal mtDNA inheritance: a crustacean, an amphibian, and a mammal.

Our initial objective was to apply the method of Awadalla, Eyre-Walker, and Maynard Smith (1999) to cases drawn from the literature. We used the study by Meyran, Gielly, and Taberlet (1998) of a segment of cytochrome oxidase I in *Gammarus fossarum* as one such case. In the process of analyzing these data, we realized that the signature of recombination was clearly preserved in the amino acid sequences. As a result, we turned our search to studies that reported coding mtDNA sequences with substantial differences at the amino acid level, even if the sequences were drawn from different species of a genus. We confined the search to within volume 16 (year 2000) of *Molecular Phylogenetics and Evolution*, from which we selected the study by Martin et al. (2000) on rodents of the genus *Apodemus* and the study by Sumida, Ogata, and Nishioka (2000) on frogs of the genus *Rana*. These two studies report only parts of the known sequences of the cytochrome *b* gene, but our analysis was based on the full DNA sequences, which we retrieved from GenBank (accession numbers and references in fig. 1).

The protocol of the analysis was as follows:

1. DNA sequences were translated into amino acid sequences (the *Drosophila* code was used for *Gammarus*, and the mammalian code was used for *Rana* and *Apodemus*). Sequences that differed at the nucleotide level but had the same amino acid sequence were treated as one sequence. Amino acid sites that either were invariable or contained a different amino acid residue in only one sequence were removed.
2. The resulting sequences were visually inspected for the identification of "polymorphic domains," i.e.,

strings of sites occupied by alternative blocks of amino acid residues that were sufficiently similar among some sequences, yet different among others (fig. 1).

3. We returned to nucleotide sequences and chose three sequences from each of the three species groups for analysis at the DNA sequence level. The triad was selected so that two sequences, designated sequence A and sequence C, could be treated as "parental" sequences, and the third, designated sequence B, could be treated as "recombinant" (fig. 2A). This designation is arbitrary given that it is not possible to separate the available sequences (fig. 1) into those that were originally parental and those that resulted from recombination. Sequence B shared a middle region (region 2) with high similarity to sequence C. We refer to this region as the "recombinant" part. Sequence B also shared a high degree of similarity with sequence A at the two regions that flanked the recombinant part (regions 1 and 3). Jointly, these two regions will be referred to as the "nonrecombinant" part.
4. The putative points of recombination were determined by visual inspection of the three DNA sequences. The two-parameter K (Kimura 1980) and the one-parameter synonymous (K_S) and nonsynonymous (K_A) divergences (Jukes and Cantor 1969) were determined for each of the three regions into which the sequences were divided (fig. 2B). We defined "postrecombination divergence" as the weighted sum of three divergence values: the divergence of the recombinant part of sequence B from the homologous part of sequence C, and the divergence of the two nonrecombinant parts of sequence B from the homologous parts of sequence A.
5. To find out if there were other ways of dividing the sequences into three regions that could produce a lower postrecombination divergence, we performed the "sliding-window" experiment. Starting from the beginning of the aligned sequences, a length equal to the recombinant part was determined and treated as a "putative" recombination region. The postrecombination divergence of the sequences so divided was determined. Then, the window was moved 10 nt downstream, and the new postrecombination divergence was determined. This process was continued until the whole length of the sequences (which, for this purpose, were treated as circular molecules) was covered. The postrecombination divergence was plotted against the nucleotide position at the beginning of the sliding window. Proper determination of the recombination fragment predicts that the smallest postrecombination divergence will occur when the sliding window coincides with the predetermined recombinant fragment. The experiment was repeated with a window length that was longer by 20 nt, and

Key words: mtDNA recombination, COI, *cyt b*, *Gammarus*, *Rana*, *Apodemus*.

Address for correspondence and reprints: Eleftherios Zouros, Institute of Marine Biology of Crete, P.O. Box 2214, 71003 Iraklio Crete, Greece. E-mail: zouros@imbc.gr.

Mol. Biol. Evol. 18(11):2127–2131, 2001

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

	111111	Accession
	223344668899900001	number
	64713281236679024681	
Gf5	I I S L S M I D T G V T M P T T M F I T	X97168
Gf1	I I S L S M I D T G I T M P T T M F I T	X97164
Gf15	V I S L S M I D T G I T L T N T M L V I	X97178
Gf17	V I T M S V I E P G I A L T N T M F V I	X97180
Gf20	V I T M S V I D T G I A L T N T M F V I	X97183
Gf14	V I T M S V I E P G I A L T K T M F V I	X97177
Gf13	I I T M S V I D T G I A L T N T M F V T	X97176
Gf11	V L S V A V V D T A V T P V A T L V V	X97174
Gf12	V L S I A V V I D A V T P V A T L V V	X97175
Gf10	V L S L E L V D T G V T L P L T L F L V	X97173
	1111222223333333	
	122362599233470223566	
	5746072935746258144069	
Rp7	I G S A A I F P A M Y V M S F I I T A T L S	AF205088
Rn8	I G T A A V F P A M Y V M S F I I T A T L S	AF205087
Rc6	M G S A A I F S A M Y V M S F I I T A T L S	AF205089
Ra1	M Y T S A I F S V L Y V M S L I I T A T L S	AF205094
Rd5	M G T S A I F F L M Y V M S F T I T A T L S	AF205090
Rd4	M G T S A A I F M F I L I F I A A I S I T	AF205091
Rr2	I V T T S V L S T L F I L I L T T A I S I T	AF205093
Rr3	M V T T S V L S T L Y V M S F I T A I S I T	AF205092
	111222222222333333	
	13448019333333459034777	
	19262880234567675048456	
Aa1	L I M L M T I A V L M M I S F M V I V I D K M	AB032854
As26	L I T L M T I A V L M M I S F S M I V I D K M	AF160603
Af2	L I M L L T I A V L M M A S F S M I V I D K M	AB032853
As24	L I M L M I V A V L M M V S F M I V I I D N L	AB033695
Ap5	L V I V M T I A I F I L V G S M I I I V D K I	AB032850
As25	I I I V M T I T I F M M I M F T I I I V N K M	AB033694
Aa7	F I M I L T I A I F M L I L F M I I V V D K M	AB032848
As6	F I V I M M I T I L I M I M F T I V V V D K M	AB032849
Aa4	F L V L M A V A I V I M I M F T V I V V N N M	AB032851
Ag3	L L M I L M V T V L M M I L S M I I I D K L	AB032852

FIG. 1.—Amino acid sequences of three studied species groups: *Gammarus* (G), *Rana* (R), and *Apodemus* (A). Numbers indicate the amino acid positions in the complete amino acid sequence. Amino acid groups in frame indicate domains that are occupied by alternative consensus sequences of amino acid residues. Different consensus sequences of the same domain are shown in different shades. Bold letters indicate residues that are unique in the consensus sequence. Species designations are as follows: Gf, *Gammarus fossarum* (Meyran, Gielly, and Taberlet 1998); Rr, *Rana rugosa* (H.-I. Lee et al. 1999); Rp, *Rana plancyi*; Rn, *Rana nigromaculata*; Rc, *Rana catesbeiana*; Ra, *Rana amurensis* (J.-E. Lee et al. 1999); Rd, *Rana dybowskii* (Kim et al. 1999); Aa1, *Apodemus alpicola*; Aa4, *Apodemus agrarius*; Aa7, *Apodemus argenteus*; As6, *Apodemus speciosus*; Ap, *Apodemus peninsulae*; Ag, *Apodemus gorkha*; Af, *Apodemus flavicollis* (Serizawa, Suzuki, and Tsuchiya 2000); As24, *Apodemus sylvaticus*; As25, *Apodemus semotus* (Suzuki, Tsuchiya, and Takezaki 2000); As26, *Apodemus sylvaticus* (Jansa, Goodman, and Tucker 1999).

then again with a window that was shorter by 20 nt. Proper determination of the length of the recombinant fragment predicts that up word or down word changes of the window size will result in a larger postrecombination divergence at the point at which the window coincides with the actual recombinant fragment. This is because a larger window will increase the divergence between sequences B and C at the recombinant region, and a smaller window will increase the divergence between sequences B and A at the flanking regions (fig. 2C).

- The next step was the “permutation” experiment. The purpose of this experiment was to see if random shuffling of nucleotide sites could generate a “recombinant part” and a “nonrecombinant part” of the same lengths as the predetermined recombinant and nonrecombinant regions (step 3), with a postrecombination divergence lower than the one produced by the presumed recombination. Let n be the length of the three compared sequences and k be the length of the exchanged region ($k < n$). The test proceeded as follows: A nucleotide site was chosen at random, and the corresponding nucleotides of the three sequences filled in the first position of the computer-generated recombinant region. The second position was filled in by a random drawing from the remaining $n - 1$ sites. This process was repeated k times, i.e., until the recombinant part was completed. The remaining $n - k$ sites were also arranged at random to produce the nonrecombinant part. The experiment was repeated 1,000 times, and the postrecombination divergence was calculated each time.
- The division of the sequences into recombinant and nonrecombinant regions implies that nucleotide divergence between sequences A and C (nonrecombinant sequences) should be the same along the entire length. However, regional divergences between sequences A and B should fall into two groups: those of the two flanking regions should be smaller than those of the middle region. This should also apply for sequences B and C, but in the opposite way. The regional homogeneity of Kimura’s K values was examined by dividing the three regions in lengths of 25 or 50 bp and testing the means by ANOVA.

The *Gammarus* comparison involved a segment of 366 bp of the cytochrome oxidase I (COI) mtDNA gene. The 20 available sequences were reduced to 10 according to the protocol of this study (step 1). Twenty of the 122 amino acid sites were variable among the 10 sequences examined (fig. 1). Two domains, the first of seven amino acids and the second of nine amino acids, were identified, each with two alternative consensus sequences. Within each consensus, block homogeneity was high, with only three sequences carrying unique replacements. Three sequences, Gf11, Gf12, and Gf10 (apparently related to each other) had diverged from the rest to the point where the domains could not be recognized. Sequence Gf15 presents clear evidence for recombination. The results shown in figure 2 extend this

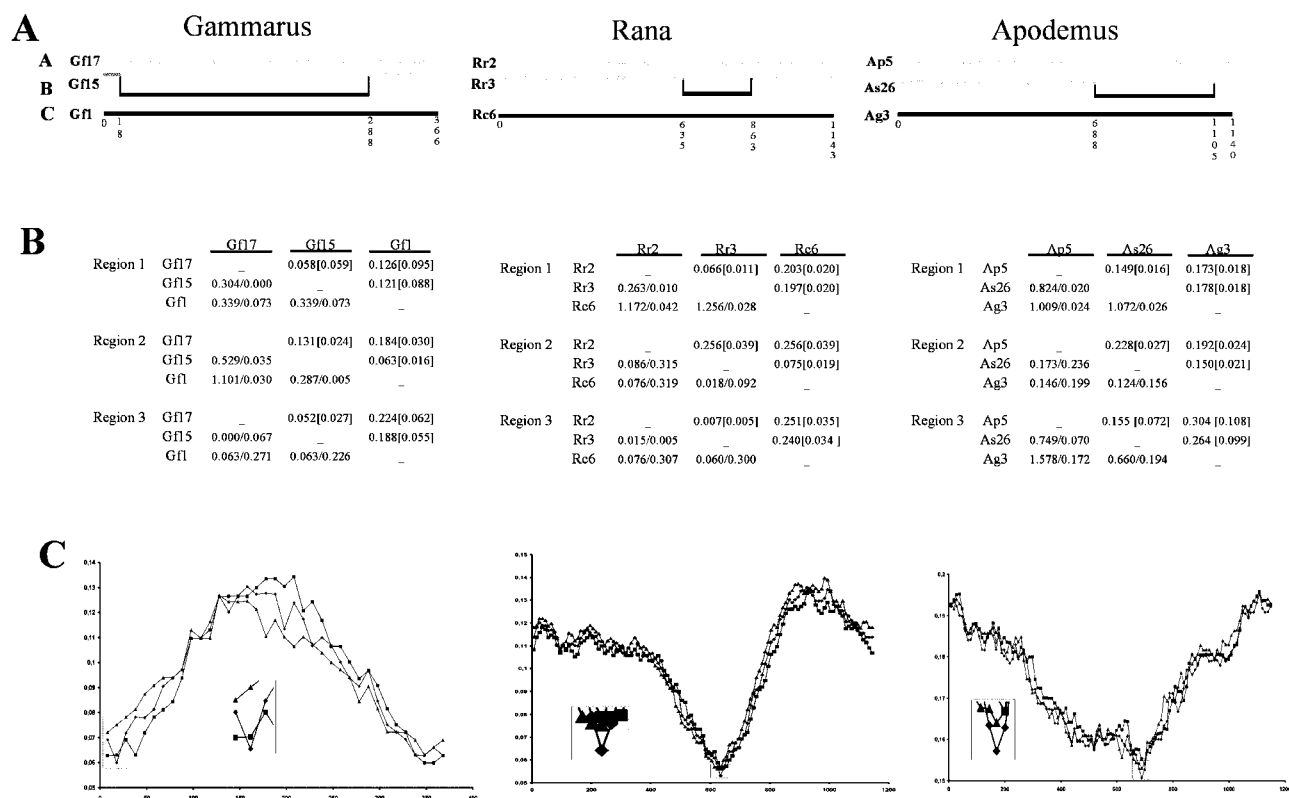


FIG. 2.—Evidence for recombination from DNA sequences. A, Presentation of sequences according to presumed recombination. Sequences A, B, and C and regions 1, 2, and 3 are defined in the text. Sequence length and putative points of recombination are shown on sequence C. Regions of similar sequence are drawn in same shade and closer to each other. B, Divergence between regions of the three nucleotide sequences. Above the diagonal: the Kimura (1980) two-parameter divergence (with SE in parentheses); below the diagonal: the synonymous (K_S) (first number) and nonsynonymous, (K_A) (second number) one-parameter Jukes and Cantor (1969) divergence. C, Results from the “sliding-window” experiment. Abscissa: nucleotide site number; ordinate: postrecombination divergence. Insert: enlarged part around the point of recombination. Diamonds: sliding window equal to the middle region; triangles: longer by 20 nt; squares: shorter by 20 nt. Note that the postrecombination divergence is smallest when the window coincides with the postulated recombination region and is of the same length.

evidence to the DNA level. The recombinant fragment was defined as the region between sites 19 and 288. This divides sequence Gf15 into three parts: one (the middle region) with high similarity to Gf1 ($K = 0.063$ for 269 bp), and two (flanking regions) with high similarity to Gf17 ($K = 0.058$ for 18 bp and 0.052 for 79 bp). This gives a postrecombination divergence of 0.060. No other continuous fragment could give a better postrecombination divergence, and no random assemblage of nucleotide sites out of 1,000 gave a lower value. The test for regional homogeneity was conducted by dividing the sequences into segments of 25 nt. As expected, there was no regional heterogeneity between the two nonrecombinant sequences Gf17 and Gf1 (from ANOVA for the means of the three regions, $F_{(2,12)} = 0.78$, $P = 0.47$), but there was significant heterogeneity between Gf15 and Gf1 ($P = 0.03$). The analogous test for Gf15 and Gf17 was not, however, significant ($P = 0.16$). This disagreement with expectation is most likely due to the short length of the tested DNA regions. The postrecombination divergence compared with the divergence between the two nonrecombinant sequences could provide an estimate of the time of the event of recombination. Given that the divergence between Gf1 and Gf17 is 0.190 and the postrecombination divergence is 0.060, we estimate that the recombination event is about one

third as recent as the time of the divergence of Gf17 and Gf1.

The *Rana* comparison involved a segment of 1,143 bp of the cytochrome *b* gene. The 14 available sequences were reduced to eight for our analysis. Twenty-two of the 381 amino acid sites were variable, and three domains were identified: one of nine, one of four, and one of six amino acid sites (fig. 1). Replacements within consensus groups were rare, as in the case of *Gammarus*. Exchanges were seen among all three domains (first with second, second with third, third with first). For DNA comparison, we chose sequences Rr2, Rr3, and Rc6. The recombinant fragment was defined as the region between nucleotide sites 636 and 863 (length 228 bp), with a corresponding postrecombination divergence of 0.053. No other continuous or random assemblage of nucleotide sites (steps 5 and 6 of the protocol) could produce a smaller value. For the homogeneity test, the sequences were divided into lengths of 50 nt. All three comparisons produced the expected results (for Rr2 vs. Rc6, $F_{(2,21)}$, $P = 0.241$; for Rr2 vs. Rr3, $P = 0.000$; for Rr3 vs. Rc6, $P = 0.005$). The divergence between Rr2 and Rc6 is 0.225. Given that the postrecombination divergence is 0.053, the time of the recombination event must be a quarter as recent as the time since the divergence of Rr2 and Rc6.

The *Apodemus* comparison involved a segment of 1,140 bp of the cytochrome *b* gene. All 10 original sequences were retained for the analysis. Twenty-three of the 380 amino acid sites were variable. We could identify two domains, one of five and one of six amino acid sites. One difference from *Gammarus* and *Rana* was that no domain could be identified in the last part of eight amino acid sites (or more than a third of the DNA sequence). Several combinations of three sequences can be used to trace the recombination at the DNA level. We used sequences Ag3, Ap5, and As26 as a triad that illustrates the recombination most clearly. The recombination fragment was defined as the region between sites 688 and 1105. Again, there was no continuous part of the DNA sequence that could produce a smaller post-recombination divergence, and only one out of 1,000 random combinations of nucleotide sites produced a lower value than the postulated recombination. As in the case of *Rana*, the sequences were divided into lengths of 50 nt for the homogeneity test. Again, all three comparisons produced the expected results (for Ap5 vs. Ap3, $F_{(2,21)}$, $P = 0.244$; for Ap5 vs. As26, $P = 0.001$; for As26 vs. Ap3, $P = 0.040$). The divergence between Ag3 and Ap5 is 0.184, and the postrecombination divergence is 0.149, which suggests that the recombination is 80% as old as the separation of the parental sequences.

The observation that recognizable amino acid domains with alternative consensus sequences occur in different combinations in a collection of sequences is in itself strong evidence for recombination. One could invoke some kind of within-domain convergent evolution to account for the observations, for example, that the first amino acid replacement in a domain may act as an agent for selective replacements in other sites. This type of selection, of interest in itself, cannot account for the nucleotide similarities. In particular, it cannot explain the parallel changes in amino acid similarities and similarities in synonymous sites (fig. 2B). To reinforce this point and to make alternative hypotheses such as bias of codon usage less likely, we removed from each triad of sequences the codons responsible for variable amino acids and calculated nucleotide divergence on the truncated sequences. In all cases but one, the results were similar to those shown in the upper part of figure 2B; i.e., for both the first and the second flanking regions, the smallest divergence was between sequences A and B, and for the middle region the smallest divergence was between sequences B and C. The exception was in the third region of the *Apodemus* triad, where the largest divergence occurred between As26 and Ap5. This exception is most likely due to the old age of the recombination event, which allowed for the accumulation of a large degree of divergence between shared sequence parts.

We conclude that amino acids simply provide stronger signals than nucleotides themselves of a process that causes regional homogenization of divergent DNA sequences. This process must be reciprocal homologous recombination, and not gene conversion alone. Evidence for this hypothesis comes from the case of *Rana* and *Apodemus*, in which all four combinations

of consensus sequences for the first two amino acid domains can be found in the collection of sequences of figure 1 (e.g., Rd5, Rd4, Rr2, and Rr3 in *Rana* and As24, Ap5, As6, and Ag3 in *Apodemus*). The difference between the postrecombination divergence and the two nonrecombinant sequences (sequence A vs. sequence C; fig. 2A) provides an estimate of the degree of divergence between sequences at the time of recombination. This degree is 0.13 for *Gammarus*, 0.03 for *Apodemus*, and 0.17 for *Rana*. Recombination between sequences with a higher degree of divergence would probably be unlikely to occur (Rayssiguier, Thayer, and Radman 1989), whereas recombination between sequences with a lower degree would be difficult to detect.

The direct evidence of homologous recombination we have reported in mussels (Ladoukakis and Zouros 2001) and the three cases of historical recombination we report here in three more species, each from a different major division of the animal kingdom, leave little doubt that recombination occurs in animal mtDNA. That we were able to come across three cases of recombination in the literature with not much searching effort suggests that many mtDNA lineages may contain evidence for recombination. This does not imply that it would be easy to estimate the frequency of mtDNA recombination. Extant recombinant sequences must represent only a small fraction of sequences generated by recombination. Yet, our observations help to reinforce the point that mtDNA genomes may recombine freely. This rarely results in new haplotypes owing to the limited frequency of heteroplasmy, which in turn is due to the rarity of biparental inheritance of animal mtDNA. The rate of paternal mtDNA "leakage" in animals has not been studied in a systematic way. The only crude estimates we have at present are 10^{-4} from laboratory crosses in mice (Gyllensten et al. 1991) and 10^{-3} per fertilization in *Drosophila* hybrids (Kondo et al. 1990; Kondo, Matsuura, and Chigusa 1992). Finally, our results provide support for the claim of recombination in hominid mtDNA (Awadalla, Eyre-Walker, and Maynard Smith 1999; Eyre-Walker, Smith, and Maynard Smith 1999), with whatever consequences this may have on biohistorical inferences based on this genome.

Acknowledgments

We thank Drs. I. Karakassis and N. Primikiriou for help.

LITERATURE CITED

- AWADALLA, P., A. EYRE-WALKER, and J. MAYNARD SMITH. 1999. Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* **286**:2524–2525.
- EYRE-WALKER, A., N. H. SMITH, and J. MAYNARD SMITH. 1999. How clonal are human mitochondria? *Proc. R. Soc. Lond. B Biol. Sci.* **266**:477–483.
- GYLLENSTEN, U., D. WHARTON, A. JOSEFSSON, and A. C. WILSON. 1991. Paternal inheritance of mitochondrial DNA in mice. *Nature* **352**:255–257.
- JANSA, S. A., S. M. GOODMAN, and P. K. TUCKER. 1999. Molecular phylogeny and biogeography of Madagascar's native

- rodents (Muridae: Nesomyinae): a test of the single origin hypothesis *Cladistics* **15**:253–270.
- JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp 21–132 in H. N. MUNRO, ed. *Mammalian protein metabolism*. Academic Press, New York.
- KIM, Y.-R., D.-E. YANG, H. LEE, J.-E. LEE, H.-I. LEE, S.-Y. YANG, and H.-Y. LEE. 1999. Genetic differentiation of mitochondrial cytochrome b gene of the Korean *Rana dybowskii* (Amphibia: Ranidae). *Korean J. Biol. Sci.* **3**:199–205.
- 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.
- KONDO, R., E. T. MATSUURA, and S. I. CHIGUSA. 1992. Further observation of paternal transmission of *Drosophila* mitochondrial DNA by PCR selective amplification method. *Genet. Res.* **59**:81–84.
- KONDO, R., Y. SATTA, E. T. MATSUURA, H. ISHIWA, N. TAKAHATA, and S. I. CHIGUSA. 1990. Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. *Genetics* **126**:657–663.
- LADOUKAKIS, E. D., and E. ZOUROS. 2001. Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. *Mol. Biol. Evol.* **18**:1168–1175.
- LEE, H.-I., D.-E. YANG, Y.-R. KIM, H. LEE, J.-E. LEE, S.-Y. YANG, and H.-Y. LEE. 1999. Genetic variation of the mitochondrial cytochrome b sequence in Korean *Rana rugosa* (Amphibia; Ranidae). *Korean J. Biol. Sci.* **3**:89–96.
- LEE, J.-E., D.-E. YANG, Y.-R. KIM, H. LEE, H.-I. LEE, S.-Y. YANG, and H.-Y. LEE. 1999. Genetic relationships of *Rana amurensis* based on mitochondrial cytochrome b gene sequences. *Korean J. Biol. Sci.* **3**:303–309.
- MARTIN, Y., G. GERLACH, C. SCHLOTTERER, and A. MEYER. 2000. Molecular phylogeny of European murid rodents based on complete cytochrome b sequences. *Mol. Phylogenet. Evol.* **16**:37–47.
- MEYRAN, J. C., L. GIELLY, and P. TABERLET. 1998. Environmental calcium and mitochondrial DNA polymorphism among local populations of *Gammarus fossarum* (Crustacea, Amphipoda). *Mol. Ecol.* **7**:1391–1400.
- RAYSSIGUIER, C., D. S. THAYLER, and M. RADMAN. 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* disrupted in mismatch-repair mutants. *Nature* **342**:396–401.
- SERIZAWA, K., H. SUZUKI, and K. TSUCHIYA. 2000. A phylogenetic view on species radiation in *Apodemus* inferred from variation of nuclear and mitochondrial genes. *Biochem. Genet.* **38**:27–40.
- SUMIDA, S., M. OGATA, and M. NISHIOKA. 2000. Molecular phylogenetic relationships of pond frogs distributed in the Palearctic region inferred from DNA sequences of mitochondrial 12S ribosomal RNA and cytochrome b genes. *Mol. Phylogenet. Evol.* **16**:278–285.
- SUZUKI, H., K. TSUCHIYA, and N. TAKEZAKI. 2000. A molecular phylogenetic framework for the Ryukyu endemic rodents *Tokudaia osimensis* and *Diplothrix legata*. *Mol. Phylogenet. Evol.* **15**:15–24.
- ZOUROS, E., A. O. BALL, C. SAAVEDRA, and K. R. FREEMAN. 1994. Mitochondrial DNA inheritance. *Nature* **368**:818.

FUMIO TAJIMA, reviewing editor

Accepted July 13, 2001