Letter to the Editor

Unequal Synonymous Substitution Rates Within and Between Two Protein-Coding Mitochondrial Genes

Joseph P. Bielawski and John R. Gold

Center for Biosystematics and Biodiversity, Department of Wildlife and Fisheries Sciences, Texas A&M University

Genes encoded by mitochondrial DNA display considerable heterogeneity of sequence change. Among protein-coding genes, rates and patterns of nonsynonymous nucleotide substitutions are influenced by differential selective constraints relative to amino acid composition and function of the gene product. Synonymous substitutions are presumably free of functional constraints on amino acid sequence. However, patterns of synonymous substitutions are not uniform. Observations of transition bias at synonymous sites across many groups of closely related taxa have led to general agreement that transitional substitutions, as a class, evolve more rapidly than transversional substitutions. Furthermore, strong nucleotide compositional bias present at largely synonymous third codon positions suggests that even within transitions or transversions the pattern of nucleotide substitution may not be uniform (Perna and Kocher 1995). Substitutional processes of mitochondrial genes are complex and additionally can vary with taxonomic group.

In this study, we examined nucleotide sequence evolution of two protein-coding mtDNA genes (cytochrome *b* and ND4L) among closely related species of the North American cyprinid genus *Notropis*. Cytochrome *b* (cyt *b*) is a well-characterized gene whose predicted amino acid sequence appears to be conserved (Irwin, Kocher, and Wilson 1991; Martin and Palumbi 1993; Cantatore et al. 1994). ND4L is a smaller and less well-characterized gene whose predicted amino acid sequence is thought to be much less conserved than that of cyt *b* (Johansen, Guddal, and Johansen 1993; Meyer 1993).

A 533-base-pair (bp) fragment of the cyt *b* gene and the entire ND4L gene (297 bp) were sequenced from representatives of eight species of the cyprinid genus *Notropis*. These included six species of the subgenus *Notropis* (*N. amabilis, N. atherinoides, N. oxyrhynchus, N. shumardi, N. stilbius,* and *N. telescopus*), *N. rubellus* (placed in the subgenus *Notropis* by Mayden and Matson [1988]), and *N. girardi* (placed in the subgenus *Notropis* by Coburn and Cavender [1992]). Species were sampled from wild populations by seining.

Genomic DNA was prepared using phenol : chloroform extraction and selected genes were amplified using the polymerase chain reaction (PCR). The PCR thermal profile consisted of 35 cycles of 95°C denaturation phase for 1 min, 48–50°C annealing phase for 1 min, and 72°C extension phase for 45 sec. Double-stranded DNA amplification products were sequenced directly using a modification of a standard cycle-sequencing protocol. The cycle-sequencing thermal profile consisted of 30 cycles of 95°C denaturing for 30 sec, 60°C annealing for 30 sec, and 72°C extension for 1 min. Primers for cyt *b* were: L 15344, 5′-ATACATGCCAACGGA-GCATC-3′; and H 16002, 5′-TCCTCGTTG-TTTTGAGGTGTG-3′. Primers for ND4L were: L 10421, 5′-CAAGACCCTTGATTTCGGCTCA-3′; and H 12293, 5′-CAAGAGTTTCAGGCTAAGACCA-3′.

Estimates of pairwise sequence divergence between taxa were computed for each gene and for each codon position, using the computer program MEGA (Kumar, Tamura, and Nei 1993). In addition, measures of the frequency of synonymous substitution were estimated for all taxa. The frequency of synonymous substitution per synonymous site was computed according to the method of Nei and Gojobori (1986), as implemented in the computer program MEGA (Kumar, Tamura, and Nei 1993). Alternatively, the frequency of synonymous pyrimidine (T-C) substitutions per synonymous T-C site and synonymous purine (A-G) substitutions per synonymous A-G site were estimated separately. Synonymous substitutions at first codon positions were excluded to facilitate computation. The numbers of synonymous A-G and T-C sites were computed between each pair of sequences, and then the frequency of synonymous substitution was computed from the respective sites for all pairwise sequence comparisons. Genetic distance was estimated using the two-parameter model of Kimura (1980). The level of substitutional saturation of each gene was assessed by graphic evaluation of transitional change at third positions versus genetic distance. The Kolmogorov-Smirnov two-sample, one-tailed test was used to determine whether pairwise estimates of sequence divergence computed using one gene were significantly larger than values computed from the other gene. The Kolmogorov-Smirnov one-sample test was used to determine whether pairwise estimates of sequence divergence for each gene were consistent with a random sample.

Differences in nucleotide composition and/or substitutional saturation can influence comparisons of sequence divergence among taxa for the two genes. Measures of nucleotide composition in both genes indicate a similar bias against guanine. Guanine composition is approximately 17% in both ND4L and cyt *b*. Among third codon positions, guanine composition is 11% in ND4L and 13% in cyt *b*. This bias was not unexpected as a similar bias has been detected in other species of fish (Meyer 1993). The strong nucleotide bias at third positions suggests that all nucleotide substitutions may

Key words: synonymous substitution, mitochondrial DNA, cytochrome *b*, ND4L, *Notropis*.

Address for correspondence and reprints: Joseph P. Bielawski, Center for Biosystematics and Biodiversity, Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, Texas 77843-2258. E-mail: j-bielawski@wfscgate.tamu.edu.

Mol. Biol. Evol. 13(6):889–892. 1996

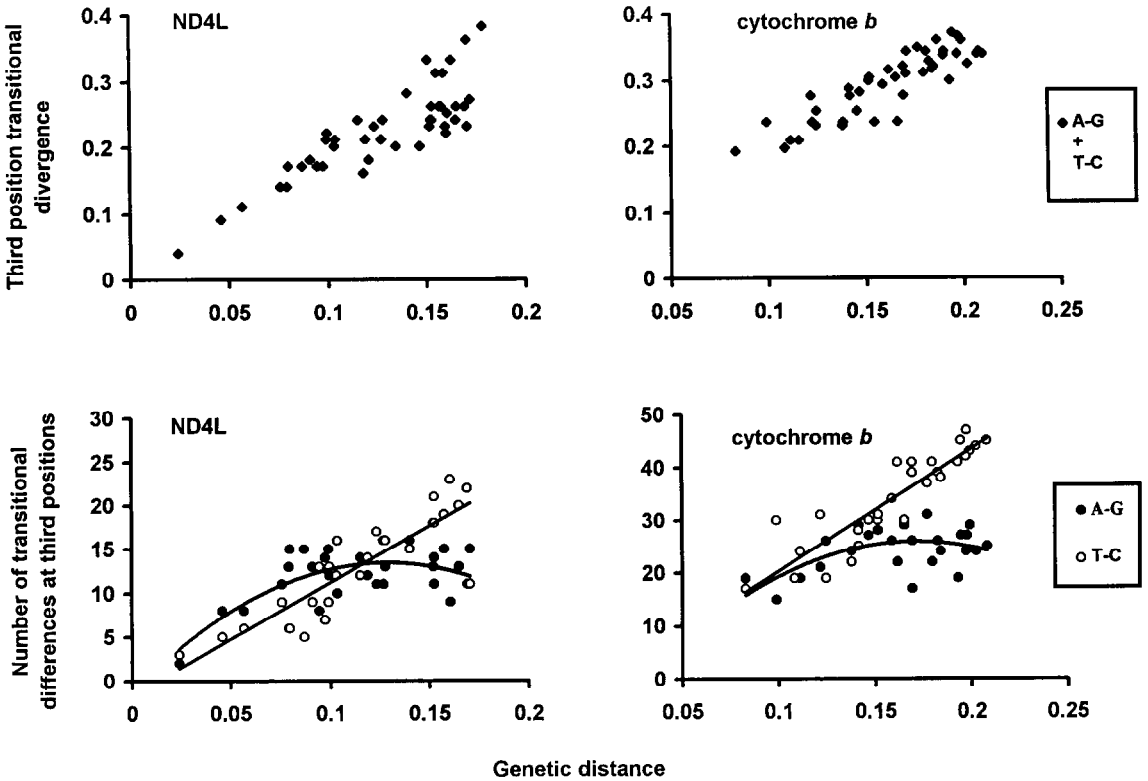

FIG. 1.—Plots of total third-position transitional divergence and plots of A-G and T-C transitions at third positions relative to Kimura's (1980) two-parameter genetic distance. Trendlines were fit to data series using a binomial model for A-G transitions and a linear model for T-C transitions.

not occur with the same frequency (Perna and Kocher 1995). This could result in a lower saturation ceiling for certain classes of nucleotide substitutions (Kondo et al. 1993). Graphic measures of total transitional divergence at third positions suggest that neither gene is saturated (fig. 1). However, separate plots of A-G and T-C transitional divergence at third positions indicate that A-G transitions are saturated in both genes, whereas T-C transitions are not (fig. 1). These data suggest that A-G and T-C transitional classes are evolving differently among these species of *Notropis*, and that measures of total transitional substitution at third positions can fail to detect these differences.

Average sequence divergence among the species of *Notropis* was significantly lower in ND4L as compared to cyt *b* (table 1). Analysis of each codon position revealed that divergence at third positions also was significantly lower in ND4L, whereas divergence at first and second positions was slightly higher in ND4L (table 1). On the whole, these data suggest that ND4L is evolving more slowly than cyt *b* among these species. However, sequence divergence is a reflection of two different processes: mutation rate, and selective constraints acting on the protein product of a gene. Because T-C transitional change at third positions appears not to be saturated in either gene (fig.

Table 1
Summary of Nucleotide Substitution Statistics for ND4L and Cytochrome *b* (cyt *b*)

	ND4L	cyt *b*	Kolmogorov-Smirnov Two-Sample Test
Mean sequence divergence	10.0%	13.8%	$K_D = 13$, $P < 0.01$
Mean sequence divergence at first positions	4.8%	3.6%	$K_D = 13$, $P < 0.01$
Mean sequence divergence at second positions	0.5%	0.1%	$K_D = 12$, $P = 0.01$
Mean sequence divergence at third positions	25.0%	38.2%	$K_D = 18$, $P < 0.01$
Mean frequency of synonymous T-C substitutions per synonymous T-C site[a]	0.176	0.305	$K_D = 15$, $P < 0.01$
Mean frequency of synonymous A-G substitutions per synonymous A-G site[a]	0.249	0.340	$K_D = 17$, $P < 0.01$
Mean frequency of synonymous substitutions per synonymous site[b]	0.355	0.522	$K_D = 16$, $P < 0.01$

[a] First-position synonymous substitutions were excluded from this measure of synonymous substitutions.
[b] This measure of synonymous substitutions was computed according to the method of Nei and Gojobori (1986).

1), the frequency of synonymous T-C substitutions per synonymous T-C site should be the best measure of absolute substitution rates in these two genes. Among the species examined, the frequency of synonymous T-C substitutions per synonymous T-C site was significantly lower in ND4L than in cyt *b* (table 1). Interestingly, the equivalent measure of the synonymous A-G substitution rate also indicated a difference between the genes; however, this measure suggests a lower difference between the substitution rates of the two genes (table 1). The latter was not unexpected given the evidence for saturation of A-G transitions (fig. 1). The Nei and Gojobori (1986) measure of synonymous substitution, an alternate measure of substitution rate that includes all pathways of synonymous substitutions between each pair of codons, also was significantly lower in ND4L than in cyt *b* (table 1). Kolmogorov-Smirnov one-sample tests of T-C divergence at third positions and of the Nei and Gojobori (1986) measure of synonymous substitutions indicate that, for each gene, lineage-specific divergence values were consistent with a random sample (data not shown). This suggests that lineage-specific effects are not a factor in comparisons of these statistics. Collectively, these data indicate that the absolute substitution rate in ND4L is lower than in cyt *b*, a finding at odds with the concept that ND4L is the more rapidly evolving gene (Johansen, Guddal, and Johansen 1993; Meyer 1993).

In addition to greater sequence divergence at first and second codon positions within ND4L (table 1), a greater percentage of amino acid positions are variable within ND4L (7%) as compared to cyt *b* (4.5%). Assuming the absolute substitution rate to be lower in ND4L, these contrasting patterns suggest that a higher proportion of nucleotide sequence positions may be free to vary in ND4L. This could explain how, for deeper phylogenetic comparisons, ND4L appears to be evolving more rapidly than cyt *b* (Johansen, Guddal, and Johansen 1993; Meyer 1993).

The apparent difference in absolute substitution rate between ND4L and cyt *b* may be related to their location in the mitochondrial genome. Mitochondrial genome replication is an asymmetric process initiated at two different times from two different origins of replication (Clayton 1982). A consequence of this mode of replication is a period where a region of light strand DNA is single-stranded and presumably more susceptible to damage by oxidative radicals. Genes close to the origin of heavy-strand replication are presumably in the single-strand state for longer periods during replication, and it has been hypothesized that these genes will have the highest mutation rates (Brown and Simpson 1982; Tanaka and Ozawa 1994). Cyt *b* is located adjacent to the origin of heavy-strand replication, whereas ND4L is located approximately halfway between the heavy- and light-strand origins of replication. Consequently, cyt *b* enters into the single-strand state first and could be subjected to oxidative damage for a longer period than ND4L. This could explain the observed higher absolute substitution rate in cyt *b* than in ND4L.

Complete nucleotide sequence from a number of mitochondrial genes spanning the region from the origin of heavy-strand replication to the origin of light-strand replication is needed to test the hypothesis that the observed difference in absolute substitution rates of ND4L and cyt *b* in these species of *Notropis* is related to genome location. A larger data set would allow examination of evolutionary differences among classes of nucleotide substitution within and between genes. We are now studying the effects of unequal nucleotide substitution rates on phylogeny reconstruction and examining patterns of nucleotide substitution inferred from phylogenetic hypotheses of these closely related species.

Acknowledgments

We thank L. R. Richardson for laboratory expertise, K. A. Dunn for helpful discussions, R. L. Honeycutt for constructive comments on an early draft, and two anonymous reviewers for their comments. Funding was provided by the National Science Foundation (Award BSR-90-20217) and the Texas Agricultural Experimental Station (Project H-6703). This paper is contribution No. 47 of the Center for Biosystematics and Biodiversity at Texas A&M University.

LITERATURE CITED

BROWN, G. G., and M. V. SIMPSON. 1982. Novel features of animal mitochondrial DNA evolution as shown by sequences of rat cytochrome oxidase subunit II genes. Proc. Natl. Acad. Sci. USA **79**:3246–3250.

CLAYTON, D. A. 1982. Replication of animal mitochondrial DNA. Cell **28**:693–705.

CANTATORE, P., M. ROBERTI, G. PESOLE, A. LUDOVICO, F. MILELLA, M. N. GADALETA, and C. SACCONE. 1994. Evolutionary analysis of cytochrome *b* sequences in some perciformes: evidence for a slower rate of evolution than in mammals. J. Mol. Evol. **39**:589–597.

COBURN, M. M., and M. T. CAVENDER. 1992. Interrelationships of North American cyprinid fishes. Pp. 328–373 *in* R. L. MAYDEN, ed. Systematics, historical ecology, and North American freshwater fishes. Stanford Univ. Press, Stanford, Calif.

IRWIN, D. M., T. D. KOCHER, and A. C. WILSON. 1991. Evolution of the cytochrome *b* gene of mammals. J. Mol. Evol. **32**:128–144.

JOHANSEN, S., P. H. GUDDAL, and T. JOHANSEN. 1990. Organization of the mitochondrial genome of the Atlantic cod, *Gadus morhua*. Nucleic Acids Res. **18**:411–419.

KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

KONDO, R., S. HORAI, Y. SATTA, and N. TAKAHATA. 1993. Evolution of hominid mitochondrial DNA with special reference to the silent substitution rate over the genome. J. Mol. Evol. **36**:517–531.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: molecular evolutionary genetics analysis, version 1.02. Inst. Mol. Evol. Gen., Pennsylvania State Univ., University Park, Pa.

MARTIN, A. P., and S. R. PALUMBI. 1993. Protein evolution in different cellular environments: cytochrome *b* in sharks and mammals. Mol. Biol. Evol. **10**:873–891.

MAYDEN, R. L., and R. H. MATSON. 1988. Evolutionary relationships of eastern North American cyprinids: an allozyme

perspective. Abstr. 86th Annual Meeting of Amer. Soc. Ichthyol. and Herpetol., Ann Arbor, Mich.

MEYER, A. 1993. Evolution of mitochondrial DNA in fishes. Pp. 1–38 *in* P. W. HOCHACHKA and P. MOMMSEN, eds. Biochemistry and molecular biology of fishes. Vol 2. Elsevier Press, The Netherlands.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

PERNA, N. T., and T. D. KOCHER. 1995. Unequal base frequencies and the estimation of substitution rates. Mol. Biol. Evol. **12**:359–361.

TANAKA, M., and T. OZAWA. 1994. Strand asymmetry in human mitochondrial DNA mutations. Genomics **22**:327–335.

FRED W. ALLENDORF, reviewing editor

Accepted February 29, 1996