Sequence Evolution and Phylogenetic Signal in Control-Region and Cytochrome b Sequences of Rainbow Fishes (Melanotaeniidae)

Danqing Zhu, Barrie G. M. Jamieson, Andrew Hugall, and Craig Moritz
Department of Zoology, The University of Queensland

The nucleotide sequences of segments of the cytochrome b gene (351 bp), the tRNA^{Pro} gene (49 bp), and the control region (~313 bp) of mitochondrial DNA were obtained from 26 fish representing different populations and species of *Melanotaenia* and one species of *Glossolepis*, freshwater rainbow fishes confined to Australia and New Guinea. The purpose was to investigate relative rates and patterns of sequence evolution. Overall levels of divergence were similar for the cytochrome b and tRNA control-region sequences, both ranging from <1% within subspecies to 15%–19% between genera. However, the patterns of sequence evolution differed. For the cytochrome b gene, transitions consistently exceeded transversions, the bias ranging from 4.2:1 to 2:1, depending on the level of sequence divergence. However, in the control-region sequence, a bias toward transitions (2:1) was observed only in comparisons between very similar sequences, and transversions outnumbered transitions in comparisons of divergent sequences. Graphic comparisons suggested that the control region was saturated for transitions at relatively low levels of sequence divergence but accumulated transversions at a greater rate than did the cytochrome b sequence. These distinct patterns of base substitution are associated with differences in A+T content, which is 70% for the tRNA control-region segment versus 50% for cytochrome b. A test for skewness in the distribution of lengths of random trees indicated that both segments contained phylogenetic signal. Parsimony analyses of the data from the two regions, with or without weighting schemes appropriate to the respective patterns of sequence evolution, identified the same five groupings of sequences, but the relationships among the groups differed. However, in most cases the branches uniting different combinations of groups were poorly supported, and the differences among topologies were insignificant. Considering the observed patterns of base substitution and the results of the phylogenetic analyses, we deduce that both the control region and cytochrome b are appropriate for population genetic studies but that the control region is less effective than cytochrome b for resolving relationships among divergent lineages of rainbow fishes.

Introduction

The development of direct sequencing using the polymerase chain reaction (PCR) has led to widespread and rapidly increasing use of mitochondrial DNA (mtDNA) sequence information over a broad range of taxonomic levels, from populations to those among phyla. This breadth of applications is made possible by differences in evolutionary rates among different mtDNA genes and among different positions within genes (reviewed by Brown 1985; Kondo et al. 1993). At the extremes, the genes encoding ribosomal RNA and some proteins evolve relatively slowly, enabling distant comparisons and resolution of deep phylogenetic branches (e.g., see Meyer and Wilson 1990; Ballard et al. 1992), whereas sections of the noncoding control region (CR) in vertebrates evolve rapidly and are suitable for resolving relatively recent relationships (e.g., see Edwards 1993).

The phylogenetic utility of gene sequences has been augmented by considering patterns of nucleotide substitution and weighting characters as appropriate (e.g., see Irwin et al. 1991; Moritz et al. 1992). A common observation has been that transitions (TS) greatly outnumber transversions (TV) among closely related mtDNA sequences, the bias often exceeding 10:1 (Brown et al. 1982). This bias is reduced between more distantly related sequences, presumably because of saturation of TS and obliteration of TS by TV (e.g., see DeSalle et al. 1987). However, there have been reports of differences among genes in the extent of the bias (Kocher and Wilson 1991; Pumo et al. 1992).

Recent studies of poikilothermic vertebrates have revealed differences relative to homeotherms in the dynamics and patterns of mtDNA evolution that could

Key words: mitochondrial DNA, sequence evolution, phylogenetic analysis, Melanotaeniidae.
Address for correspondence and reprints: Craig Moritz, Department of Zoology, The University of Queensland, Brisbane QLD 4072 Australia.

Mol. Biol. Evol. 11(4):672–683. 1994.
© 1994 by The University of Chicago. All rights reserved.
0737-4038/94/1104-0012$02.00

well affect the way in which the molecule is used in systematics. Poikilotherms appear to have a substantially slower rate of nucleotide substitution (Martin and Palumbi 1993), and the rate of amino acid substitution appears to be slower in salmonids (Thomas and Beckenbach 1989) and other cold-blooded vertebrates (Adachi et al. 1993). Also, length variation within and around the noncoding region appears to be more common in cold-blooded than in warm blooded vertebrates (reviewed by Moritz et al. 1987).

Several studies of protein-coding and tRNA sequences in fish mtDNA (Kocher et al. 1989; Thomas and Beckenbach 1989; Beckenbach et al. 1990; Meyer et al. 1990; Digby et al. 1992) have reported a strong bias toward TS, as in mammalian sequences. In contrast, there have been several indications that the CR in fish mtDNA may differ from that of mammals, in having less of a TS bias or in having a substitution rate similar to that of protein-coding regions (Johansen et al. 1990; Bernatchez et al. 1992; Fajen and Breden 1992; Shedlock et al. 1992; but see Sturmbauer and Meyer 1992). However, none of these studies included matched comparisons of protein-coding versus CR sequences over a range of divergences sufficiently wide to allow a proper comparison of the dynamics of sequence evolution. The present study aims to redress this omission by comparing cytochrome b (cytb) and CR sequences from different populations and species of rainbow fishes of the genus *Melanotaenia.*

The family Melanotaeniidae consists of eight genera, some termed "rainbow fishes," including the genus *Melanotaenia,* which contains 32 species or subspecies that are confined to New Guinea and Australia (Allen 1989). Much recent research has been done on morphological classification in *Melanotaenia* (Allen and Cross 1982; Crowley et al. 1986; Allen 1989). However, there is both considerable between-species overlap in morphometric characters and some uncertainty about species boundaries. We have sampled fish of the genus *Melanotaenia* from different taxonomic levels and from widely separated geographic areas, to provide a broad spectrum of mtDNA divergences. This sampling design assumes that the taxonomic hierarchy is roughly correlated with evolutionary divergence but not that current species boundaries or species groupings are necessarily accurate.

Material and Methods

Sample Collection

Fish samples were collected from Queensland, Western Australia, South Australia, the Northern Territory, and Papua New Guinea. The localities are listed in the legend to figure 1. For the same species, different sampling locations were selected. Some individuals were obtained from captive stocks where wild-caught material was not available. Where possible, two individuals of each taxon were used for analysis, according to the principle that it is desirable to replicate at the level below that of interest (Smouse et al. 1991). The keys and descriptions of Allen and Cross (1982) were used to identify the specimens, with the aid of fish taxonomists (see Acknowledgments).

DNA Extraction

In the initial stages, mtDNA was extracted and purified from some specimens by ultracentrifugation (Dowling et al. 1990). For routine PCR amplifications, crude preparations of DNA were obtained from 1 mg of fresh or frozen muscle tissue by lysing cells in the presence of 500 μl 5% Chelex (Bio-Rad) at 56°C overnight. Samples were then heated at 95°C–100°C for 15 min, vortexed, and briefly centrifuged; 2 μl of supernatant was used per 12.5-μl PCR reaction.

PCR Amplification and Sequencing

All sequences were obtained by direct sequencing of PCR products either by the asymmetric method (Gyllensten and Erlich 1988) or by cycle sequencing using the BRL Thermal Cycle Dideoxy DNA Sequencing Kit. Primers were designed from preliminary rainbow fish sequences obtained using "universal" primers (Kocher et al. 1989). For the cytb region the primers designed for rainbow fishes were L 14859, 5′-AACTTTGGCTCCCTACTTGG-3′; and H 15322, 5′-GTGGAKGATAAGRARGTGCT-3′; for the tRNA^{Pro} gene+CR we used L 15957, 5′-TTAAATTCCCTCCTAATGCTATC-3′; and H 16449, 5′-CACGATATTTGTCCCTGACC-3′ ("L" and "H" refer to the light and heavy strands, respectively, and the number refers to the position of the 3′ base in human mtDNA sequence [Anderson et al. 1981]; K=½G+½T; and R=½A+½G).

Sequences were obtained via asymmetric amplification as described by Moritz et al. (1992), with annealing at 52°C for 1 min and extension at 72°C for 1 min. Negative controls were always included. For cycle sequencing the product from a 25-μl PCR reaction was gel-purified with the Prep-A-Gene DNA Purification Kit (Bio-Rad) to remove unincorporated primers and dNTPs. About 20 ng of PCR product was sequenced from both directions by using the BRL cycle sequencing kit with primers end-labeled with (γ^{33}P-dATP) by polynucleotide kinase. Except for a few light-strand sequences obtained by asymmetric amplification, sequences were determined in both directions.

```
M.s.spl.I    CTTATTTATCACATGCTCCCTCCCTAACCCACCACCCACCCAAGCACACAACCTTCCATTATTCCTACCGTAAGCTACCTCATGAACCCTCCCA
M.s.spl.I    ..C...C...................................................................................?..
M.s.spl.II   ..C...C.........A...C......................................C................................T.
M.s.spl.II   ..C...C.........A.T.C...C....................G.............C................................T.
M.s.rub.     ..C...C...........T.C......................................C............T..........A....S....?
M.s.ino.     ...C.....T........T.C......................................C..................................
M.s.her.     ..ACC.C..T....A.A.T.C.....G....GT..TTG.TG....T.....T.C.....GC.....T..TC.CC.C.TT...CA.R..A....C
M.s.her.     ..ACC.C..T....A.A.T.C.....G....GT..TTG.TG....T.....T.C.....GC.....T..TC.CC.C.TT...CA.R..A....C
M.s.aus.I    ..CTC.CG....G.AT..T.......GTT...TG.....T....T..GT.G..C.....CG.C.T...A...C.C......C..T..T.TT.C
M.s.aus.I    ..CTC..G....G.AT..T.......GTT...TG.....T....T..GT.G..C.....CG.C.T...A...C.C......C..T..T.T..C
M.s.aus.II   ..GCC.C..TG...A...T.C..........AT.....TTT.GA.G.....T.....TC.G..TT.G......A.C..T..GCA.G...CT.TC
M.s.aus.II   ..GCC.C..TG...A...T.C..........AT.....TTT.GA.G.....T.....TC.G..TT.G......A.C..T..GCA.G...CT.TC
M.eac.I      ..CCC..G....G.ATATT..........T...TG.....TT...A..G..G..C.....C..C.T.G..A...C.C......CA.G....TT.T
M.eac.II     ..CCC..G....G.AT.TT...T.....T...TG.....TT...A..G..G..C.....C..C.T.G..A...C.C......CA.G....TT.C
M.eac.II     ..CCC..G....G.AT.TT...T.....T...TG.....TT...A..G..G..C.....C..C.T.G..A...C.C......CA.G....TT.C
M.dub.       A.CCC.C.......AT..T..........T..GYG.....TT...G...............C....T.C.........GA.....A.R....T?.T
M.dub.       .CGCC.C........T.............T..GYG......T...A...............C....T.C.........GA.......R...C...T
M.nig.       TCAC....C.....A...T.C.T........GT.....TTT..A.C............CCC......GT..C..A.C..T...CA.G..TCT..C
M.mac.       ..C.C.C.C.........T.C..........A........T.................C.............A......T.CA......TT..
M.mac.       ..C.C.C...........T.C..........A........T.................C.............A......T.CA......TT..
M.tri.       ..ACC.C..T....A.A.T.C.....G....GTG.T.G.TGG...T.....T.C...C.CGC....T..TC..C.C.TT...CA.G...C.?.C
M.exq.       ..GCC.C..TG..CA...A.C..........AT.....TTT.GA.G...........TC.G...T.G..C...A.C..T..GCA.G..TCT.TC
M.exq.       ..GCC.C..TG..CA...A.C..........AT.....TTT.GA.G...........TC.G...T.G..C...A.C..T..GCA.G..TCT.TC
M.par.       ....C.C...........T.C...C...................A..............C.............A.C.T.....A......T..C
M.aff.       T.ACC.C.C.....A.A.TACT.TCT.TTTG.T....G.T....T.T....T.CCG.C.CG..T.CTTTACCTCACG..G.CC..T..T....C
G.inc.       T.ATCCC.C..T....A.TACT..CT.T....T.TT.T.T.GCA.G.G.G.TG..ATCCGGC...CT..ACTCAAC....T.CAGGGT.C...C
             3912233344555566778889911111111111111111111111111222222222222222222222222222223333333333333
              81736928124769251476900001112222223444566777788888900011112344555666677778888999000012223334
                                          2568147034692147658124703469814703698469258145703692568158036981473695
```

FIG. 1.—Aligned sequence data showing only variable nucleotide positions for 351 bp of cytb, corresponding to positions 14927–15279 in human mtDNA for 26 taxa examined, consisting of 13 species and subspecies of *Melanotaenia* and one species of *Glossolepis*. Roman numbers refer to different populations. The abbreviations of taxa are as follows: M.s.spl. = *Melanotaenia splendida splendida* (I—Davies Creek, North Queensland; and II—Burdekin River, North Queensland); M.s.rub. = *M. s. rubrostriata* (New Guinea species, locally bred, Victoria); M.s.ino. = *M. s. inornata* (Cohen River, North Queensland); M.s.her. = *M. s. herbertaxelrodi* (New Guinea species locally bred, Victoria); M.s.aus. = *M. s. australis* (I—Bellary Creek, Western Australia; and II—South Alligator River, Northern Territory); M.eac. = *M. eachamensis* (I—Hanson stock; and II—Walkamin Station); M.dub. = *M. duboulayi* (Delanip Creek, Woodford, Queensland); M.nig. = *M. nigrans* (Jardine River, Cape York Peninsula, North Queensland); M.mac. = *M. maccullochi* (Jardine River, Cape York Peninsula, North Queensland); M.tri. = *M. trifasciata* (Gunshot Creek, Cape York Peninsula, North Queensland); M.exq. = *M. exquisita* (South Alligator River, Northern Territory); M.par. = *M. parkinsoni* (New Guinea species locally bred, Victoria); M.aff. = *M. affinis* (Buvu Creek, Morobe Papua, New Guinea); and G.inc. = *Glossolepis incisus* (Lake Sentani, Papua New Guinea). The sequence for *M. s. splendida* is shown in the first line, and only nucleotides different from those in this sequence are indicated in the subsequent sequences. ? = Uncertain; Y = C or T; S = C or G; and R = A or G. GenBank accession numbers are X77504 and X77507.

Sequence Alignment and Analysis

The cytb sequences were aligned, by eye, with published sequences from cichlids (Meyer et al. 1990) and sturgeon (Brown et al. 1989). The CR sequences were aligned using Clustal V with various input orders, a gap penalty of 10, and TV weighted over TS by 2:1. Changing the input order did not affect the alignment produced. The resulting alignments were inspected and adjusted manually to improve local matches. Regions that remained ambiguous were excluded from analysis (see below). Observed numbers of nucleotide substitutions, TS, and TV were compiled using "Macsequence," a program in BASIC written by D. Good. Sequence divergences were calculated using the Kimura two-parameter model provided in Clustal V. Phylogenetic relationships were estimated using PAUP (version 3.0s; Swofford 1991) and the neighbor-joining method in PHYLIP. MacClade 3.0 (Maddison and Maddison 1992) was used to explore tree lengths of alternative topologies.

Results

Sequence Alignment and Divergence

Sequences were obtained from two regions of the mtDNA molecule: (1) a 351-bp segment of the 5′ end of the cytb gene (fig. 1) and (2) a continuous sequence comprising the 3′ 49 bp of the proline transfer RNA ($tRNA^{Pro}$) gene and the following 313 bp of the noncoding CR (fig. 2). For the cytb sequence, 94 of the 351 positions were variable among the 26 different sequences, with 76 positions being informative for parsimony analysis. Most (85 of 94) substitutions occurred in third-base codon positions, the remainder being in the first codon position. Only three replacement substitutions were observed. There were no deletions or insertions, and all sequences could be translated according to the standard vertebrate mtDNA code. The average A+T content was 53% overall and 50% in the third codon position. As in other vertebrates (Brown 1985), the proportion of G at the third codon position, 5%, was very low.

<-tRNA^{Pro} |

```
M.s.spl.I   AGACTGCATCCCG---------CGAGATATATA--CCATTTGCCACATAGCTTACATTAAAAATTATATTTAAACTTAC
M.s.spl.I   ...........-----------...........--......A.....A.A.............................
M.s.spl.II  ...........-----------.......A...--......A.......A.............................
M.s.spl.II  ...........-----------...........--......A.......A...................C.........
M.s.rub.    ...........-----------..A........--......A.......A...................C....A....
M.s.ino.    ........C..-----------...........--......A.......A..........G..................
M.s.her.    GA....-....-----------T.GAC..--..TATA....T.....-GA..C......G..TA.G.T..---------
M.s.her.    GA....-....-----------T.GAC..--..TATA....T.....-GA..C......G..TA.G.T..---------
M.s.aus.I   ...G...-C..-----------..GA....-..TA-.....A.......A...........GT.A.CT..A.G......
M.s.aus.I   G..G...-C..-----------T.GA....-..TA-.....T.......A...........GT.A.CT..A.G......
M.s.aus.II  .......-C..-----------...........--......A....GG.ASK..A....G..T.......A.G.T....
M.s.aus.II  .......-C..-----------...........--......A....GG.A....A.A..G..T.......A.G.T....
M.eac.I     ...G...-..T.----------..GA...CTA.TA-..........CAGA..C.A..CT...T.A.CT..A.G.T....
M.eac.II    ...G...-..T.----------..GA...C..ATA-..........CA.A...GA..CT...T.A.CT..A.G.T....
M.eac.II    ...G...-..T.----------..GA...C..ATA-..........CA.A...GA..CT...T.A.CT..A.G.T....
M.dub.      .......-..T.----------..GA...A...--......A....GG.A.....R......T.A..TC.A........
M.dub.      .......-..T..---------..GG...A...--.....KA....GG.A............T.A..TC.A........
M.nig.      .....C.-C..-----------..GA.......--....A.A....GG.A......A..G..T.......A.G.T....
M.mac.      .......-...-----------......TC...--......A.......A.............................
M.mac.      ......GT...-----------......TT...--......A.......A.....R.......................
M.tri.      .......-...-----------T.GAC..C-..TATA....T.TGT.-GA..C.T..C.G..TA.GCT..---------
M.exq.      .......-C..-----------S..........--......A....GG.A......A..G..T.......A.G.T....
M.exq.      .......-C..-----------...........--......A....GG.A......A..G..T.......A.G.T....
M.par.      ...........-----------...S.......--......A...S...ASK...........................
M.aff.      .......-..G.AAAGTTTTAA...AC.TA...CC..TC..T.....G.A............T...CC..A.G.AAC..
G.inc.      ..-.C...--T.GCAAATTTCAA-------...--..TG.AC-....CGA........T.G.T...CT..A.G..AA..
            1255556666666667777777888899999911111111111111111111111111111111111111111112222
            3634890234567890123456678901238900000000111111223333445556667788889999999990000
                                            01234568014589013568590175780801230134567890123

M.s.spl.I   ATAACTTAATATAGTACATTATTGTCAATTGAAGGCTTTATACTATCAAACCTTTAACCGAACGAA-TATCTATTAATT-AGTA
M.s.spl.I   ..................................................................-............-....
M.s.spl.II  ..................................................................-...........C-....
M.s.spl.II  ....A........AC....C......................A.......T...............-C..T.G......-.T..
M.s.rub.    ..................................................................-...........A-....
M.s.ino.    ............G.....................................................-...........C-....
M.s.her.    -------------------------------..A.....AT..-.TG........C.ATA......-C.CT..C....C-..A.
M.s.her.    -------------------------------..A.....ATT.-.TG........C.ATA......-C.CT???....C-.???
M.s.aus.I   ....TA..G.T..ACT................CA...A..TT.-C............A........-A.....C.....-GA..
M.s.aus.I   ....TA..G.T..ACT...............GTA...A.CTT.-C............A........-......C....C-GA..
M.s.aus.II  ...T.A.G...GCAC.......................T.TTC-.T..TA..C..A......T.A........C....C-.CA.
M.s.aus.II  .....A.G...GCAC.......................T.TTC-.T..TA..C..A......T.A........C....C-.CA.
M.eac.I     ....TA....T..ACT..........G.........GC..TT.-.............A........-C.....C....C-.T..
M.eac.II    ....TA....T..ACT.................A..GC..TT.-.............A........-C.....C...CC-.T..
M.eac.II    ....TA....T..ACT.................A..GC..TT.-.............A........-C.....C....C-.T..
M.dub.      ....TA....T....C...............GG.......TTC-......TTT...GA........-C...........-.A..
M.dub.      ....TA....T....C...............GG.......TTC-......TTT...GA........-C...........-.A..
M.nig.      ....AA.......TAC...............GG...AA..T..T.......T..............-...T..C....C-.T..
M.mac.      ....A........AC....C...........GA.................................-...T..C....C-.T..
M.mac.      ....A..G.....AC....C..................................C...........-..CT..C....C-.T..
M.tri.      -------------------------------..A.....ATT--.T..AT........-TA...T.-C.CT..C....C-.AA.
M.exq.      .....A.G...GCAC.......................T.TTC-.T..TA..C..A......T.A........C....C-.CA.
M.exq.      .....A.G...GTAC.......................T.TTC-.T..TA..C..A......T.A........C....C-.CA.
M.par.      ....A..G.....AC....C...........CA..................T..............-....KGC.....-....
M.aff.      ..T..A.T.ATCTCCC.........ATTA..GTA...........TA.-....C.A...C.CAA..AC-CT..C....C-.AAC
G.inc.      ....AC....TACCACA........AC.A..GTAAGC.A.CTT-.......TAA...A..G.-.CC-...G..CA?G..G.TA-
            222222222222222222222222222222222222222222222222222222222222222233333333333333333333
            000000111111111122222222223333333444555555566666777777888888889901111111122334445556
            456789012345678901234567890123489079023456812357034789012457890161234578934560372472
```

FIG. 2.—Aligned sequence data, showing only variable nucleotide positions for part of the tRNA^{Pro} gene (49 bp) and CR (313 bp) from mtDNA for 13 species and subspecies of *Melanotaenia* and one species of *Glossolepis.* The sequence corresponds to positions 15977–16339 of the human mtDNA sequence. Boldface type indicates regions of uncertain alignment, which were deleted from the phylogenetic analysis. – = Insertion or deletion; and K = G or T; all other abbreviations, including species abbreviations, are as in fig. 1. GenBank accession numbers are X77505–X77506.

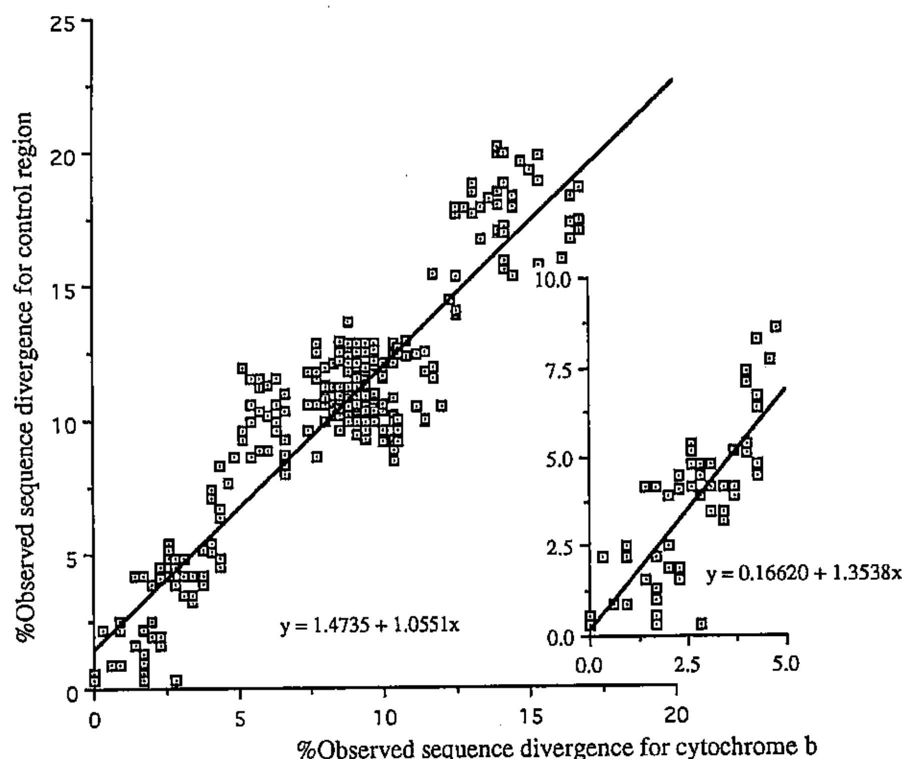

FIG. 3.—Observed sequence divergence of cytochrome b vs. control region, for pairwise comparisons encompassing two genera and 14 species and subspecies of rainbow fishes. The graph insert to the right includes just the points for which cytb sequence divergence is 0%–5% and indicates that the control region evolves slightly faster than does the cytochrome b gene when the sequence divergence is low. Note that the data points, representing all pairwise comparisons, are not independent.

In contrast to the cytb gene, the CR and $tRNA^{Pro}$ region showed a strong bias in base content, with 70% A+T. Comparisons of the alignments produced with different gap penalties and substitution weighting identified two areas of ambiguity of alignment in the CR close to the $tRNA^{Pro}$ gene (shown in boldface in fig. 2); these were excluded from further analyses. Numerous deletions or insertions, of 1–40 bp, were observed both within and between genera. The largest deletion, of 40 bp, was shared by three mtDNAs, two from *Melanotaenia splendida herbertaxelrodi* and the other from *M. trifasciata.* Considering just base substitutions, 115 of the 313 CR positions varied with 89 informative sites. The proportion of variable sites in the CR sequence (115/313) is bigger than that in the cytb sequence (94/351).

Estimates of sequence divergence by using the Kimura two-parameter model for the cytb segment were 0.0%–21.8%. Within *Melanotaenia,* the maximum value was 17.9%, with broad overlap between the ranges of values for putative interspecific and intersubspecific comparisons. The highest sequence divergence within genera occurred between the wild-caught New Guinea species *M. affinis* and the other *Melanotaenia.* The values were 12.5%–17.7%, well within the range of sequence variation at the genus level. Divergences among the CR sequences were broadly similar to those of cytb. A graphic comparison of pairwise observed sequence divergences (fig. 3) indicates that, overall, the CR sequence is not diverging faster than the cytb gene. However, considering just the more closely related sequences, i.e., those within cytb divergences of <5%, the CR segment is evolving at ~1.4 times the rate for cytb (fig. 3).

Patterns of Sequence Evolution

Inspection of the rates of transition versus transversion revealed a marked difference between the cytb and the CR sequences. For cytb, TS outnumbered TV at all levels of divergence, with values ranging from 4.2:1, between the most similar sequences, to 2.3:1, between the most divergent sequences (table 1). For the CR sequence, TS outnumbered TV in comparisons between closely related mtDNAs (as measured by cytb divergence), but between the more divergent sequences, TV were as common as or outnumbered TS (table 1). For these divergent comparisons, the CR TV also outnumbered cytb TV, by as much as 2:1, despite the two regions having similar levels of sequence divergence overall (table 1).

A plot of the percentage of TS against the percentage of TV, for all pairwise comparisons of cytb and the CR sequences (fig. 4), provides another perspective on these divergent patterns of sequence evolution. For the cytb sequence there is an approximately linear relationship between numbers of TS and TV, with a bias toward TS. In contrast, the CR sequence shows a nonlinear relationship such that TS appear to plateau at ~5%–8% as TV accumulate. This, together with the greater number of TV between divergent CR sequences (table 1 and fig. 4), suggests that TS are saturating at a low level of divergence in the CR, whereas for cytb they continue to accrue.

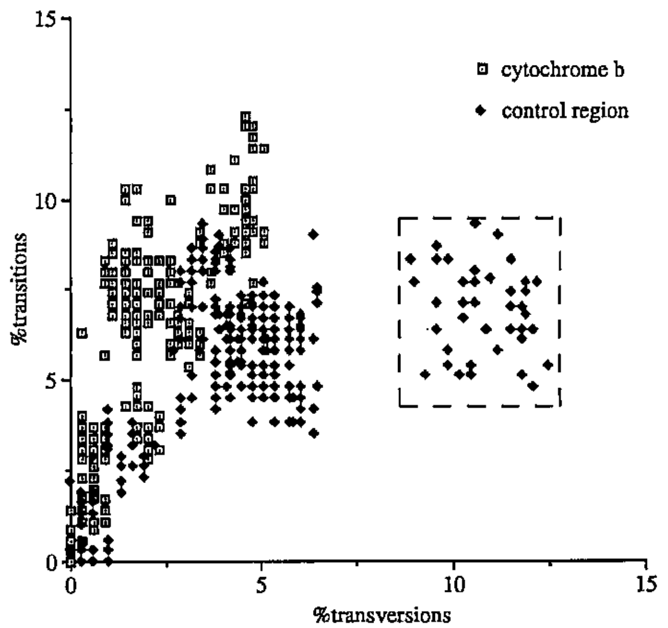

FIG. 4.—Pairwise comparisons of percent of observed sequence divergence, transversions to transitions, for cytochrome b and control-region sequences. The saturation of control-region transitions was evident in the comparison of sequence between *Melanotaenia affinis* and other *Melanotaenia* and between *Glossolepis* and *Melanotaenia*. (*boxed*)

Phylogenetic Analysis

Skewness of tree-length distributions, as a measure of information content (Hillis and Hulsenbeck 1992), was tested by generating 10,000 random trees. Both data sets produced significantly skewed distributions of random trees, with the cytb having a stronger skew ($g_1 = -0.32$; $P<0.01$) and thus greater information content than the CR region ($g_1 = -0.24$; $P<0.01$). Combining the two sets of data increased the skewness ($g_1 = -0.496$; $P<0.01$; fig. 5) and therefore the apparent information content of the data.

We used parsimony methods in PAUP, with and without character weighting, for phylogenetic analysis. Regions of the CR sequences for which alignments were uncertain were excluded from the data matrix. For unweighted characters, heuristic searches produced 16 shortest trees for the cytb sequence, 83 shortest trees for the CR, and 11 shortest trees for the combined data (fig. 6). All analyses identified the same five groupings of mtDNAs, as shown in figure 7: one (A) containing sequences from *M. s. splendida*, *M. s. rubrostriata*, *M. s. inornata*, *M. maccullochi*, and *M. parkinsoni*; a second (B) with only two individuals of *M. duboulayi*; a third (C) consisting of *M. eachamensis* and Western Australian *M. s. australis*; a fourth (D) with *M. exquisita* and individuals of *M. s. australis* (all captive-bred from the same locality in the Northern Territory) and *M. nigrans*; and a fifth group (E) including *M. trifasciata* and *M. s. herbertaxelrodi* (e.g., see fig. 7). All analyses put group E as the sister lineage to groups A–D, but the relation-

Table 1
TS and TV among cytb and CR Sequences—and Relative Levels of Sequence Divergence of CR Sequences—among Pairwise Comparisons (*n*) of mtDNAs Classified According to Divergence in the cytb Region

	cytb DIVERGENCE[a]			
	0.0%–5.0% (n = 47)	5.1%–10.0% (n = 113)	10.0%–15% (n = 132)	>15% (n = 31)
Mean cytb divergence:	2.7%	7.9%	12.0%	20.0%
cytb TS	303	2,528	3,615	1,125
cytb TV	72	695	1,064	488
cytb TS/TV	4.2	3.6	3.4	2.3
Mean CR divergence:	3.2%	9.5%	12.0%	16.0%
CR TS	325	1,896	2,595	676
CR TV	165	1,504	2,161	1,025
CR TS/TV	2.0	1.3	1.2	0.7
Mean CR/cytb divergence	1.2	1.2	1.0	0.8

[a] Calculated by the Kimura two-parameter method.

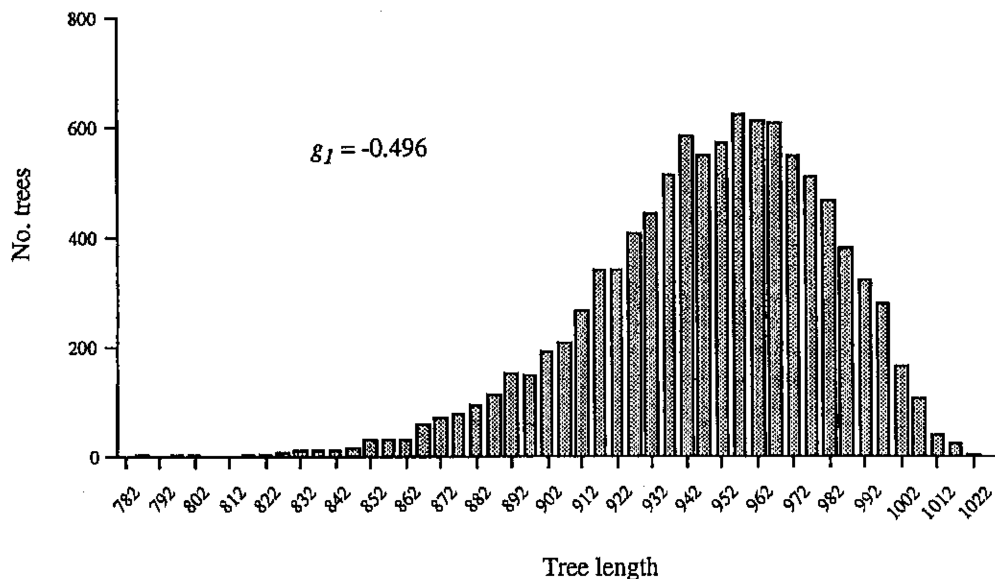

FIG. 5.—Distribution of tree lengths for 10,000 random trees from combined cytb and CR data, without weighting

ships among groups A–D differed (see fig. 6, I-1–III-1). The robustness of the branches—evaluated by bootstrapping with the membership of the groups (but not the relationship of the groups) constrained—was low, except for the branches uniting A–D to the exclusion of E.

Given (1) the consistent bias toward TS observed between similar sequences and (2) the apparent saturation of TS between even moderately divergent CR sequences, distinct weighting schemes were applied for further parsimony analyses. For cytb, TV were weighted 10 times relative to TS, whereas for the CR sequence, only TV were used. The number of shortest trees decreased to three for cytb and two for CR. The same five groupings (A–E) were present in each case, but the relationships among them again differed (fig. 6, I-2 and II-2). Combining the two data sets with their respective weighting schemes—i.e., 10:1 weighting of TV over TS for cytb and 10:0 weighting for the CR sequences—produced three shortest trees (fig. 6, III-2). As was the case for the unweighted analyses, the bootstrap values were generally low, except for that defining the A–D group in the combined analysis.

Neighbor-joining methods were also used to examine the phylogenetic relationships. The same five groupings were found by either the Jukes-Cantor or the Kimura weighting (TS:TV=10:1) model, for the combined data set. Both the Jukes-Cantor and the Kimura weighting models revealed the same phylogenetic relationships among groups as was found in PAUP analyses of the combined data set under different weighting schemes (fig. 6, III-2).

Although the consensus topologies appeared to differ between sequences and between unweighted and weighted analyses, the numbers of characters defining the relationships among groups were often low, as were the corresponding bootstrap values (figs. 6 and 7). The only branch that appeared consistent and moderately robust was that uniting the A–D groups to the exclusion of E. To further explore the differences between these topologies, the tree lengths were compared among the data sets, with and without weighting. The results (table 2) indicate that very little separates the alternative arrangements of groups A–D in each of the data sets. The exception is that the relationship between *M. affinis* and group E (trees II-1 and II-2), suggested by the CR data, appears strongly incompatible with the cytb data. Comparing the distributions of characters between the alternative topologies and the shortest tree for each type of data showed that the cytb data set, especially when weighted, was better able to discriminate than was the CR data set.

Discussion

Levels of Sequence Variation in Relation to Current Taxonomy

The levels of sequence divergence seen for the cytb and CR sequences in the rainbow fishes are within the ranges observed for some other genera of fish (e.g., *Tropheus*, ≤15%; Sturmbauer and Meyer 1992), although they are higher than in most studies (e.g., see Bernatchez et al. 1992). A crude hierarchy is apparent in relation to the existing taxonomy. Fish from different populations of the same subspecies (*Melanotaenia splendida splen-*

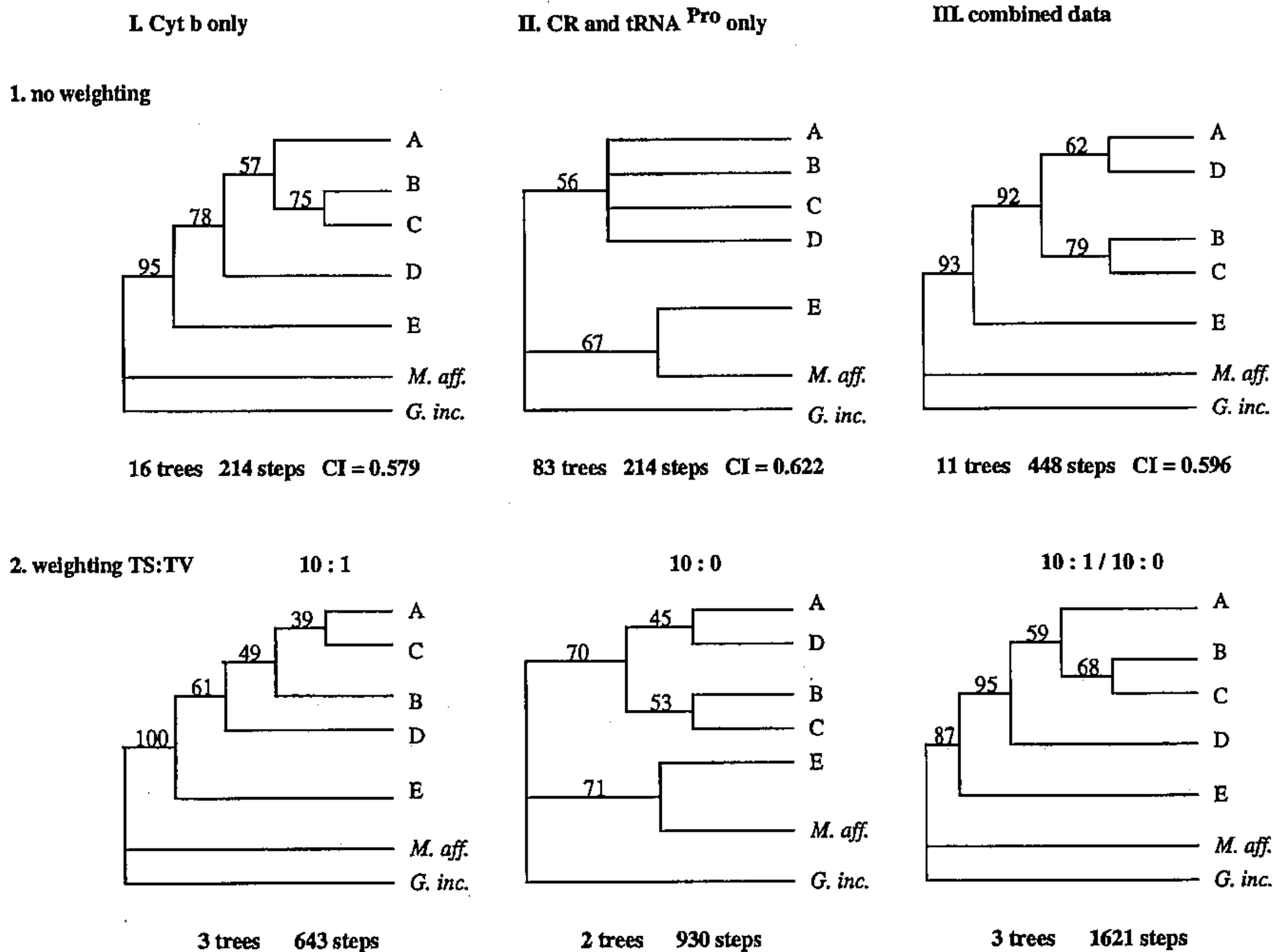

FIG. 6.—Relationships among groups of species, revealed by heuristic searches with different data sets and weighting schemes. A–E refer to the five main phylogeny groupings of sequences indicated in all analyses and shown in fig. 7. Numbers above the branches are the bootstrap values after 100 replications in heuristic searches with the membership of groups A–E constrained and *Glossolepis incisus* (*G. inc.*) as the outgroup.

dida) have <4% divergence for each sequence; fish of different genera (i.e., *Melanotaenia* vs. *Glossolepis*) have divergences of >17%; and fish of different putative subspecies—or of different species within genera—typically have intermediate values. However, there is broad overlap between the ranges of divergence for comparisons between subspecies and those between species of *Melanotaenia*, and *M. affinis* is as different from other *Melanotaenia* as is *Glossolepis*. This could reflect either errors in the current taxonomy, or sorting of ancestral polymorphisms among closely related species (Neigel and Avise 1986), or both.

The above situation is further complicated by anomalies such as the polyphyly of mtDNAs from *M. s. australis*, fish from Western Australia having mtDNAs closely related to those from *M. eachamensis*, whereas *M. s. australis* from the Northern Territory have mtDNAs very similar to those from *M. exquisita* from the same locality. Given that hybridization between different species of rainbow fishes is commonplace in captivity (Allen and Cross 1982), it is likely that the extraordinary divergence within *M. s. australis* is due to introgression from *M. exquisita*. However, the close relationship of *M. s. australis* and *M. eachamensis*, in terms of mtDNA, warrants further study.

The biological and systematic implications of these patterns will be dealt with elsewhere. The important observation here is that our sampling of mtDNAs across a range of taxonomic levels has produced a set of comparisons spanning a broad range of sequence divergences within the range typically used in population genetic and phylogenetic studies. This provides the raw material for evaluating and contrasting patterns of sequence divergence.

Comparison of Patterns of Sequence Divergence

Previous studies of mtDNA sequences from fish species have revealed conflicting patterns of base sub-

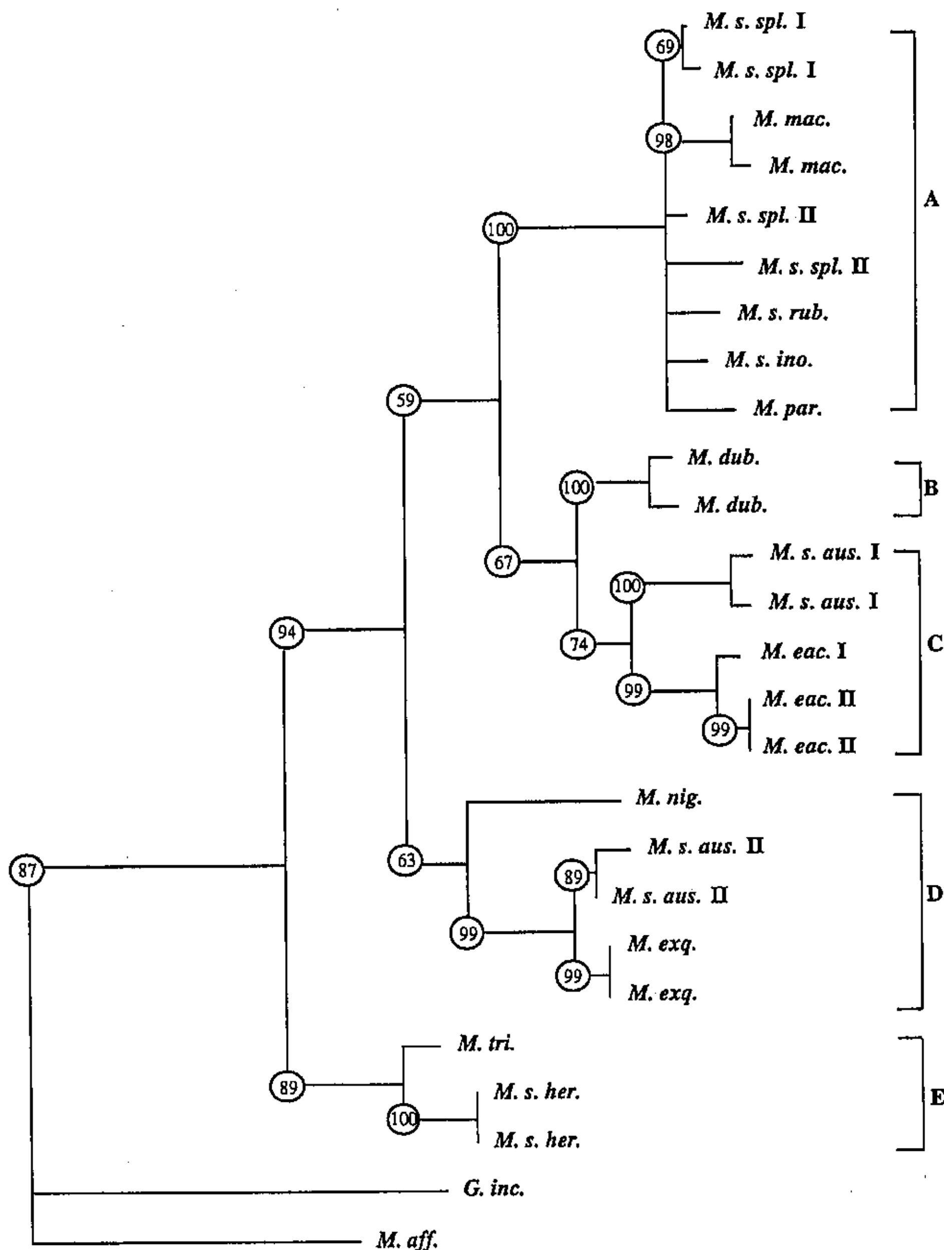

FIG. 7.—Phylogenetic tree for mtDNA sampled from 13 species of *Melanotaenia*, based on a parsimony analysis of combined cytb and CR mtDNA sequences, with weighting of TV over TS. Numbers on the branch are the bootstrap values after 100 replications with heuristic searches. The main phylogenetic groups are represented as A–E. Species abbreviations are as in fig. 1.

stitution. The bias toward TS appeared to be weak relative to that in mammals (typically, >10:1; e.g., see Brown et al. 1982) in the CRs of guppy (3:1; Fajen and Breden 1992) and various salmonid species (3:1; Bernatchez et al. 1992; Shedlock et al. 1992) and for cytb of the prickly shark (Bernardi and Powers 1992). Further, the data for salmonids suggested that most of the CR evolves at about the average rate for the genome (Bernatchez et al. 1992; Shedlock et al. 1992). However, strong biases toward TS were observed in studies of African cichlids (Sturmbauer and Meyer 1992) and sturgeon (Brown et al. 1993), and the former also showed more rapid evolution in the CR than in cytb.

The patterns observed for rainbow fishes may help to explain these contradictions, in that the most similar CR sequences show a bias toward TS, whereas sequences from more divergent mtDNAs have more TV than TS. The overall levels of sequence divergence (corrected for multiple hits, by the Kimura two-parameter model) in cytb and the CR are similar at all levels of divergence.

Table 2
Tree Lengths for Different Data Sets and Weighting Schemes Applied to the Alternative Topologies Shown in Figure 6

	ALTERNATIVE TOPOLOGY IN FIGURE 6					
DATA SET	I-1	I-2	II-1	II-2	III-1	III-2
Only cytb, no weights	214	220 (3.0%)	232 (8.4%)[a]	238 (11.2%)[a]	215 (0.5%)	220 (3.0%)[a]
Only cytb, 10:1 weighting	655 (1.9%)	643	764 (18.8%)[a]	760 (18.2%)[a]	674 (4.6%)	661 (2.8%)
Only CR, no weights	236 (2.6%)	241 (4.8%)[a]	230	233 (1.3%)	233 (1.3%)	235 (2.2%)
Only CR, 10:0 weighting	980 (5.4%)	1,000 (7.5%)	950 (2.2%)	930	970 (4.3%)	960 (3.2%)
cytb + CR, no weights	450 (0.4%)	461 (2.9%)[a]	462 (3.1%)[a]	471 (5.1%)[a]	448	455 (1.6%)
cytb + CR, 10:1/10:0 weighting	1,635 (0.9%)	1,643 (1.4%)	1,714 (5.7%)[a]	1,690 (4.3%)[a]	1,644 (1.4%)	1,621

NOTE.—The values underlined represent the shortest trees for each data set, and the percentages are the proportional increase, in tree length, over the shortest tree.

[a] Trees that are significantly poorer explanations of a data set than is its shortest tree (determined using the compare tree option in MacClade and χ^2 analysis); the tree length differences between I-1 and III-2 are due to the fine structure within the groups A–E.

However, between the more divergent sequences, there were up to twice as many TV in the CR. If TV are accumulating linearly, as appears to be the case in most comparisons of vertebrate mtDNAs (Miyamoto et al. 1989; Irwin et al. 1991), these observations suggest that the CR does have a higher evolutionary rate, but that this is masked by saturation of TS in the CR, at an unusually low level of sequence divergence.

A low "ceiling" of sequence divergence and an excess of TV due to saturation of TS has also been reported for *Drosophila* mtDNA and attributed to high A+T content (Wolstenholme and Clary 1985; DeSalle et al. 1987; Tamura 1992). Closely related mtDNAs from *Drosophila* showed a bias toward TS and rapid evolution, but the TS were saturated at ~8% sequence divergence (DeSalle et al. 1987). In rainbow fishes, the CR has a strong bias toward A+T (70%), whereas cytb has ~50% A+T in all codon positions, suggesting that apparent saturation of TS at low levels of sequence divergence in the CR is due to biased base content, as in insects (see also Crozier and Crozier 1992; Monforte et al. 1993). The same effect of compositional constraint on base substitution may operate in the salmonid CR, as the regions where the substitutions are concentrated also tend to be A+T rich (Shedlock et al. 1992).

Implications for Molecular Systematic Studies of Rainbow Fishes

In general, the evolutionary rate and pattern of base substitution in a gene are expected to affect its utility for population genetic or phylogenetic analyses (Hillis and Moritz 1990). For population genetics, the premium is on allelic and nucleotide diversity among populations of the same species. For rainbow fishes, the CR and cytb sequences appear to be equally useful, as they have similar overall rates of divergence (table 1).

For phylogenetic studies, the goal is to maximize informative variation while minimizing noise or misinformation due to homoplasy (Swofford and Olsen 1990). In this context, the apparent saturation of TS in the CR is expected to reduce the informativeness of this region, especially for resolving deeper relationships. In contrast, no such saturation was evident for cytb, although in wider intergeneric comparisons the third-base codon positions do saturate for TS (D. Zhu, unpublished data). From these observations we would expect that the cytb sequence is better able to resolve deep relationships than is the CR and that weighting should improve resolution in both cases.

In fact, the same five groups were recognized in all analyses, but none was able to clearly resolve the relationships among the groups, other than to indicate that the samples from *M. trifasciata* and *M. s. herbertaxelrodi* (group E) were consistently the most divergent. Combining the sequences from the two regions resulted in a slight increase in the stability of some branches, as revealed by bootstrapping, but may be suspect because of the heterogeneity in patterns of base substitution (Bull et al. 1993). These observations suggest that mtDNA sequence analysis is effective at identifying robust groupings within *Melanotaenia*, itself a useful contribution, but that it will be difficult to obtain a fully resolved picture of the relationships among the species groups.

Acknowledgments

We thank Rolly Mackay, Steve Donnellan, Bruce Hanson, and Andrew Tappin for providing samples; Anita Heideman for laboratory assistance; Leo Joseph for helpful discussion; and two anonymous reviewers for useful suggestions on the manuscript.

LITERATURE CITED

ADACHI, A., Y. CAO, and M. HASEGAWA. 1993. Tempo and mode of mitochondrial DNA evolution in vertebrates at

the amino acid sequence level: rapid evolution in warm-blooded vertebrates. J. Mol. Evol. **36**:270–281.

ALLEN, G. R. 1989. Freshwater fishes of Australia. T.F.H. Publications, Neptune City, N. J.

ALLEN, G. R., and N. J. CROSS. 1982. Rainbow fishes of Australia and New Guinea. Angus & Robertson. London, Sydney, Melbourne, Singapore, and Manila.

ANDERSON, S., A. T. BANKIER, G. T. BARRELL, M. H. L. DEBRUIJN, A. R. COULSON, J. DROUIN, I. C. EPERON, D. P. NIERLICH, B. A. ROE, F. SANGER, P. H. SCHVEIER, A. J. H. SMITH, R. STADEN, and I. G. YOUNG. 1981. Sequence and organization of the human mitochondrial genome. Nature **290**:457–465.

BALLARD, J. W. O., G. J. OLSEN, D. P. FAITH, W. A. ODGERS, D. M. ROWELL, and P. W. ATKINSON. 1992. Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods. Science **258**:1345–1348.

BECKENBACH, A. T., W. K. THOMAS, and H. SOHRABI. 1990. Intraspecific sequence variation in the mitochondrial genome of rainbow trout (*Oncorhynchus mykiss*). Genome **33**:13–15.

BERNARDI, G., and D. POWERS. 1992. Molecular phylogeny of the prickly shark, *Echinorhinus cookei*, based on a nuclear (18S rRNA) and a mitochondrial (cytochrome b) gene. Mol. Phylogenet. Evol. **1**:161–167.

BERNATCHEZ, L., R. GUYOMARD, and F. BONHOMME. 1992. DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. Mol. Ecol. **1**:161–173.

BROWN, J. R., A. T. BECKENBACH, and M. J. SMITH. 1993. Intraspecific DNA sequence variation of the mitochondrial control region of white sturgeon (*Acipenser transmontanus*). Mol. Biol. Evol. **10**:326–341.

BROWN, J. R., T. L. GILBERT, D. J. KOWBEL, P. J. O'HARA, N. E. BUROKER, A. T. BECKENBACH, and M. J. SMITH. 1989. Nucleotide sequence of the apocytochrome B gene in white sturgeon mitochondrial DNA. Nucleic Acids Res. **17**:4389.

BROWN, W. M. 1985. The mitochondrial genome of animals. Pp. 95–130 *in* R. J. MACINTYRE, ed. Monographs in evolution biology: molecular evolutionary genetics. Plenum, New York.

BROWN, W. M., E. M. PRAGER, A. WANG, and A. C. WILSON. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18**:225–239.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, and P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. **42**:384–397.

CROWLEY, L. E. L. M., W. IVANTSOFF, and G. R. ALLEN. 1986. Taxonomic position of two crimsonspotted rainbow fish *Melanotaenia duboulayi* and *Melanotaenia fluviatilis* (Pisces: Melanotaeniidae) from eastern Australia, with special reference to their early life-history stages. Aust. J. Mar. Freshwater Res. **37**:385–398.

CROZIER, R. H., and Y. C. CROZIER. 1992. The cytochrome *b* and ATPase genes of honeybee mitochondrial DNA. Mol. Biol. Evol. **9**:474–482.

DESALLE, R., T. FREEDMAN, E. M. PRAGER, and A. C. WILSON. 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. J. Mol. Evol. **26**:157–164.

DIGBY, T. J., M. W. GRAY, and C. B. LAZIER. 1992. Rainbow trout mitochondrial DNA: sequence and structural characteristics of the non-coding control region and flanking tRNA genes. Gene **113**:197–204.

DOWLING, T. E., C. MORITZ, and J. D. PALMER. 1990. Nucleic acids II: restriction site analysis. Pp. 250–317 *in* D. M. HILLIS and C. MORITZ, eds. Molecular systematics. Sinauer, Sunderland, Mass.

EDWARDS, S. V. 1993. Long-distance gene flow in a cooperative breeder detected in genealogies of mitochondrial DNA sequence. Proc. R. Soc. Lond. [Biol.] **252**:177–185.

FAJEN, A., and F. BREDEN. 1992. Mitochondrial DNA sequence variation among natural population of the Trinidad guppy, *Poecilia reticulata*. Evolution **46**:1457–1465.

GYLLENSTEN, U. B., and H. A. ERLICH. 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. Proc. Natl. Acad. Sci. USA **85**:7652–7656.

HILLIS, D. M., and J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. **83**:189–195.

HILLIS, D. M., and C. MORITZ. 1990. An overview of applications of molecular systematics. Pp. 502–515 *in* D. M. HILLIS and C. MORITZ, eds. Molecular systematics. Sinauer, Sunderland, Mass.

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 Atlantic cod, *Gradus morhua*. Nucleic Acids Res. **18**:411–419.

KOCHER, T. D., W. K. THOMAS, A. MEYER, S. V. EDWARDS, S. PÄÄBO, F. X. VILLABLANCA, and A. C. WILSON. 1989. Dynamics of mitochondrial DNA evolution in animals: amplication and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA **86**:6196–6200.

KOCHER, T. D., and A. C. WILSON. 1991. Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and a protein-coding region. Pp. 391–413 *in* S. OSAWA and T. HONJO, eds. Evolution of life-fossils, molecular and culture. Springer, Tokyo.

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

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution, version 3. Sinauer, Sunderland, Mass.

MARTIN, A. P., and S. R. PALUMBI. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA. **90**:4087–4091.

MEYER, A., T. D. KOCHER, P. BASASIBWAKI, and A. C. WILSON. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature **347**:550–553.

MEYER, A., and A. C. WILSON. 1990. Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. J. Mol. Evol. **31**:359–364.

MIYAMOTO, M. M., S. M. TANHAUSER, and P. LAIPIS. 1989. Systematic relationships in the artiodactyl tribe bovini (family bovidae), as determined from mitochondrial DNA sequences. Syst. Zool. **38**:342–349.

MONFORTE, A., E. BARRIO, and A. LATORRE. 1993. Characterization of the length polymorphism in the A+T-rich region of the *Drosophila obscura* group species. J. Mol. Evol. **36**:214–223.

MORITZ, C., T. E. DOWLING, and W. M. BROWN. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. **18**: 269–292.

MORITZ, C., C. J. SCHNEIDER, and D. B. WAKE. 1992. Evolutionary relationships within the *Ensatina eschscholtzii* complex confirm the ring species interpretation. Syst. Biol. **41**:273–291.

NEIGEL, J. E., and J. C. AVISE. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. Pp. 515–534 *in* S. KARLIN and E. NEVO, eds. Evolutionary processes and theory. Academic Press, New York.

PUMO, D. E., C. J. PHILLIPS, M. BARCIA, and C. MILLAN. 1992. Three pattern of mitochondrial DNA nucleotide divergence in the meadow vole, *Microtus pennsylvanicus*. J. Mol. Evol. **34**:163–174.

SHEDLOCK, A. M., J. D. PARKER, D. A. CRISPIN, T. W. PIETSCH, and G. C. BURMER. 1992. Evolution of the salmonid mitochondrial control region. Mol. Phylogenet. Evol. **1**:179–192.

SMOUSE, P. E., T. E. DOWLING, J. A. TWOREK, W. R. HOEH, and W. M. BROWN. 1991. Effects of intraspecific variation on phylogenetic inference: a likelihood analysis of mtDNA restriction site data in cyprinid fishes. Syst. Zool. **40**:393–409.

STURMBAUER, C., and A. MEYER. 1992. Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes. Nature **358**:578–581.

SWOFFORD, D. L. 1991. PAUP: phylogenetic analysis using parsimony, version 3.0. Illinois Natural History Survey, Champaign.

SWOFFORD, D. L., and G. J. OLSEN. 1990. Phylogeny reconstruction. Pp. 411–501 *in* D. M. HILLIS and C. MORITZ, eds. Molecular systematics. Sinauer, Sunderland, Mass.

TAMURA, K. 1992. The rate and pattern of nucleotide substitution in *Drosophila* mitochondrial DNA. Mol. Biol. Evol. **9**:814–825.

THOMAS, W. K., and A. T. BECKENBACH. 1989. Variation in salmonid mitochondrial DNA: evolutionary constraints and mechanisms of substitutions. J. Mol. Evol. **29**:233–245.

WOLSTENHOLME, D. R., and D. O. CLARY. 1985. Sequence evolution of *Drosophila* mitochondrial DNA. Genetics **109**: 725–744.

SIMON EASTEAL, reviewing editor

Received October 15, 1993

Accepted March 14, 1994