A MOLECULAR PHYLOGENY OF THE REMORAS AND THEIR RELATIVES

Kurtis N. Gray, Jan R. McDowell,
Bruce B. Collette, and John E. Graves

ABSTRACT

The Echeneoidea comprise three families of cosmopolitan tropical/subtropical marine fishes: the Echeneidae (remoras), Coryphaenidae (dolphinfishes), and Rachycentridae (cobia). Complete nucleotide sequences from the mitochondrial 12S rRNA, 16S rRNA, protein-coding ND2, and nuclear ITS-1 gene regions were used to reconstruct the phylogenetic history of these fishes. Parsimony, maximum likelihood, and Bayesian analyses of combined data sets resolved trees of similar topology. Congruent with evolutionary hypotheses based on larval morphology, a monophyletic Rachycentridae + Coryphaenidae was resolved with high support. Within a monophyletic Echeneidae, the subfamilies Echeneinae and Remorinae were monophyletic. In agreement with recent morphological analyses, the genus *Remora* was paraphyletic based on the position of *Remorina albescens* Temminck and Schlegel, 1850. Consistent resolution within the Remorinae using parsimony, maximum likelihood, and Bayesian inference was not achieved with the gene regions surveyed in this study.

The superfamily Echeneoidea comprises three of the 160 families that are traditionally placed in the order Perciformes: the Echeneidae, Coryphaenidae, and Rachycentridae (Nelson, 2006). The Echeneoidea has been grouped with the Carangidae (jacks, pompanos) and the Nematistiidae (roosterfish) on the basis of prenasal canal ossification and scale structure (Freihofer, 1978). Collectively, these five families make up the suborder Carangoidei (Johnson, 1984; Springer and Smith-Vaniz, 2008). The carangids and echeneoids form a monophyletic group based on three features: these fishes lack the bony stay posterior to the ultimate dorsal and anal pterygiophores, possess two prenasal canal units, and bear a lamellar expansion of the coracoid. Synapomorphies that unite the Rachycentridae, Coryphaenidae, and the Echeneidae include absence of supraneural bones, an anterior shift of the first dorsal pterygiophore, presence of several anal pterygiophores anterior to the first haemal spine, loss of the beryciform foramen in the ceratohyal, tubular ossifications surrounding both prenasal canal units, and elongate larvae with late dorsal fin completion (Johnson, 1984).

Placement of the Echeneoidea within the Carangoidei is uncontested but the phylogenetic relationships of the species within the Echeneoidea remain unresolved despite a number of morphological investigations (Freihofer, 1978; Johnson, 1984, 1993; Ditty and Shaw, 1992; Ditty et al., 1994; O'Toole, 2002). Based on larval characters relating to neurocranial development, head spination, mandibular structure, and epithelial cell composition, Johnson (1984) hypothesized a rachycentrid-coryphaenid sister group relationship. This hypothesis was supported by the work of Ditty and Shaw (1992) and Ditty et al. (1994) in an examination of larval development in cobia and dolphin, respectively. However, O'Toole (2002) published a phylogeny based on 138 putatively informative osteological characters that is inconsistent with this hypothesis. Specifically, in O'Toole's phylogeny, the Coryphaenidae were placed as

a sister group to the Rachycentridae-Echeneidae clade. In addition, the phylogeny did not support the subfamilies Echeneinae and Remorinae, or the monophyly of the genus *Remora*. This phylogeny was supported by behavioral characters concerning the development of the symbiotic "hitchhiking" and the degree of host specialization within the Echeneidae. O'Toole (2002) postulated a progression from general schooling behavior to close association with floating objects (demonstrated by the coryphaenids), which progresses further to following behavior (exhibited by *Rachycentron*), and finally, direct host attachment via a modified dorsal fin (demonstrated by the echeneids). Within the Echeneidae, a progression from coral-reef associating generalist symbiont (*Phtheirichthys, Echeneis*) to pelagic generalist [*Remora brachyptera* (Lowe, 1839)] to pelagic specialist (*Remora remora* Linnaeus, 1758) and finally, pelagic obligates [*Remora osteochir* (Cuvier, 1829), *Remora australis* (Bennett, 1840), and *Remora albescens* (Temminck and Schlegel, 1850)] was suggested (O'Toole, 2002).

Family level relationships with the Echeneoidea have also been addressed using molecular evidence. In an analysis of alpha-level taxonomy within the Carangidae, Reed et al. (2002) hypothesized a coryphaenid-rachycentrid sister-group relationship. Analyses of mitochondrial cytochrome *b* sequences resolved *Coryphaena hippurus* Linnaeus, 1758 and *Rachycentron canadum* (Linnaeus, 1766) as a monophyletic outgroup to the Carangidae using parsimony, maximum likelihood, and Bayesian inference methods. Placement of the single echeneid examined was problematic. *Echeneis naucrates* Linnaeus, 1758 was alternately resolved within and as an outgroup to the carangids studied, depending on the optimality criterion used. To date, no comprehensive molecular investigation into the taxonomic relationships within the Echeneoidea has been performed.

It is common to address phylogenetic questions using multiple gene regions (both mitochondrial and nuclear-encoded) to generate robust evolutionary hypotheses (Chen et al., 2003; Near and Cheng, 2008; Musilová et al., 2008). Because taxonomic ambiguities exist at the family, genus, and species levels within the Echeneidae, we evaluated genetic variation using four gene regions (three mitochondrial and one nuclear) that exhibit differing rates of sequence evolution. Variation was examined in the relatively slowly evolving mitochondrial 12S and 16S ribosomal RNA (rRNA) gene regions, the moderately evolving protein-coding NADH-dehydrogenase subunit 2 (ND2) gene region, and the rapidly evolving nuclear-encoded internal transcribed spacer subunit 1 (ITS-1) gene region. These gene regions have been used successfully to estimate phylogenetic relationships in a number of fish taxa (Domanico et al., 1997; Broughton and Gold, 2000; Thacker, 2003; Mattern, 2004; Westneat and Alfaro, 2005). As a comparative index of support, these data were analyzed using three different inference methods (optimality criteria): maximum parsimony (Felsenstein, 1983), maximum likelihood (Felsenstein, 1981), and Bayesian inference (Ronquist and Huelsenbeck, 2003).

In this study, complete nucleotide sequences from the four gene regions were collected from extant members of the suborder Carangoidei. Gene-based phylogenies were generated using maximum parsimony, maximum likelihood, and Bayesian inference methodologies. These phylogenies were used to test existing taxonomic hypotheses based on larval morphology (Johnson, 1984, 1993) and adult osteology and behavior characters (O'Toole, 2002). Specifically, molecular evidence was used to address the following: (1) monophyly of the superfamily Echeneoidea, (2) family level sister-group relationships within the Echeneoidea, (3) monophyly of the Echeneidae,

(4) monophyly of the subfamilies Echeneinae and Remorinae, (5) monophyly of the genus *Remora*, and (6) species-level relationships within the Echeneidae.

MATERIALS AND METHODS

To address taxonomic ambiguities within the superfamily Echeneoidea, representatives from six families of marine fishes (families Pomatomidae, Nematistiidae, Carangidae, Coryphaenidae, Rachycentridae, and Echeneidae) were collected. A single non-carangoid species, *Pomatomus saltatrix* (Linnaeus, 1766) and two non-echeneoid, carangoid species, *Nematistius pectoralis* Gill, 1862 and *Carangoides armatus* (Rüppell, 1830), were used as outgroups. With the exception of *C. armatus*, all samples were procured from coastal and offshore collections in the Atlantic and Pacific oceans. Molecular data from the complete mitochondrial sequence of *C. armatus*, Genbank accession number AP004444, supplemented the outgroup taxa data set.

Collections were made between August 2002 and July 2005 using academic, federal, commercial, and recreational fishing resources. Samples were caught by hook and line individually or in association with their pelagic hosts (e.g., billfish, sharks, rays, dolphin, and buoys). Whole specimens were placed on ice or immediately frozen to prevent tissue breakdown. Samples were identified using the keys of Lachner (1984) and Collette (2003) and photographed. Tissue samples were taken and stored in either DMSO tissue storage buffer (0.25 M disodium ethylenediamine-tetraacetatic acid (EDTA), 20% dimethyl sulfoxide (DMSO), saturated sodium chloride (NaCl), pH 8.0), or 95% ethanol. Voucher specimens are held at the Virginia Institute of Marine Science, the National Museum of Natural History, or the Scripps Institution of Oceanography.

Total genomic DNA was extracted from 0.03 to 0.10 g skeletal and/or heart muscle following the methods of Sambrook and Russell (2001). Complete double-stranded nucleotide sequences from the mitochondrial 12S, 16S, and ND2 gene regions and nuclear ITS-1 region were amplified following standard polymerase chain reaction (PCR) methodology using *Taq* PCR Core reagents (Qiagen Corp. Valencia, CA). Multiple primer sets were used to amplify gene sequences across all echeneoid samples (Table 1). Universal mitochondrial PCR primers were designed based on consensus identity of published primer sequences (Palumbi, 1996; Broughton and Gold, 2000) with gene sequences from carp, *Cyprinus carpio* Linnaeus, 1758 (Sorenson et al., 1999), little tunny, *Euthynnus alletteratus* (Rafinesque, 1810) (AB099716), and sailfish, *Istiophorus platypterus* (Shaw in Shaw and Nodder, 1792) (McDowell, 2002). Superfamily-specific internal 16S rRNA primers were designed based on consensus identity of echeneoid DNA sequences. Nuclear ITS amplifications were performed using primers designed by Johnson (2003). PCR amplification conditions consisted of an initial denaturation of 4 min at 94 °C, followed by 35 cycles of 1 min at 94 °C, 1 min at 50 °C, and 1.5 min at 72 °C, followed by a final extension of 5 min at 72 °C (with minor exceptions). Alternate cycling conditions were utilized when the above conditions were unsuccessful, and consisted of a "touchdown" cycle defined by an initial denaturation of 4 min at 94 °C, followed by 45 cycles of 1 min at 94 °C, 1 min at 56 °C (decreasing 2 °C every 5 cycles), and 1.5 min at 72 °C, followed by a final extension of 5 min at 72 °C. All PCR amplifications were performed using an MJ Research PTC-200 thermocycler (Watertown, MA). Products were electrophoresed through an agarose gel matrix, stained with ethidium bromide, and visualized using an ultraviolet-light transilluminator.

PCR products were either purified using QIAquick PCR Purification reagents (Qiagen Corp., Valencia, CA), or EXOSAP (USB Scientific, Cleveland, OH) and sequenced directly. All nuclear ITS-1 amplicons as well as mitochondrial fragments that could not be successfully direct-sequenced were cloned into a plasmid vector using the TOPO-TA plasmid cloning system (Invitrogen Corp., San Diego, CA). Cloned fragments were isolated and purified using QIAprep Spin Miniprep reagents (Qiagen Corp., Valencia, CA) following the manufacturer's specifications prior to sequencing.

Table 1. PCR and cycle sequencing primer sequences used to amplify mitochondrial 12S rRNA, 16S rRNA, ND2, and nuclear ITS-1 DNA sequences in the Echeneoidea.

Gene region	Primer	Sequence	Reference
12S rRNA	Phe-5F	5′ aaagcataacactgaagatgt 3′	This study
	Phe-5.1F	5′ aaagcrtaacactgaagatgt 3′	This study
	16S-3R	5′ accagctatmacyaggttcg 3′	This study
	16S-3.1R	5′ accagctatsacyaggttcg 3′	This study
	12SA-L (12SA-5′)	5′ aaactgggattagatacccactat 3′	Modified from Palumbi (1996)
	16SA-H (16SA-3′)	5′ atgtttttgataaacaggcg 3′	Modified from Palumbi (1996)
16S rRNA	Val-5F	5′ gcawagcatytcmcttacacyg 3′	This study
	Val-5.1F	5′ gcrtytcccttacacygagaagtc 3′	This study
	Leu-3R	5′ rytgggragaggayttgaacc 3′	This study
	Leu-3.1R	5′ rytgggagaggayttgaacc 3′	This study
	16S-IAF	5′ agttartcaaarggggkacagc 3′	This study
	16S-IBR	5′ caartgattacgctacctthgc 3′	This study
ND2	ND2B-LF	5′ taagctttygggcccatac 3′	Modified from Broughton and Gold (2000)
	ND2B-HR	5′ crrttaggrctttgaaggc 3′	Modified from Broughton and Gold (2000)
ITS-1	X18SF	5′ cttgactatctagaggaagt 3′	Johnson (2003)
	X28SR	5′ atatgcttaaattcagcggg 3′	Johnson (2003)
	5.8SR1	5′ attcacattagttctcgcagcta 3′	Johnson (2003)
	5.8SR2	5′ attgatcatcgacmyttcgaacgcac 3′	Johnson (2003)

Purified PCR products and recombinant plasmids were sequenced in forward and reverse directions following the dideoxynucleotide chain termination sequencing methodology (Sanger et al., 1977). Samples were sequenced using BigDye Terminator 3.1 Cycle Sequencing (Applied Biosystems, Warrington, UK) and electrophoresed using either an ABI 3100 or ABI 3130 DNA sequencer. Sequences were analyzed using Sequencing Analysis v. 5.1.1 software (Applied Biosystems, Warrington, UK). Alternately, samples were sequenced using Thermo Sequenase cycle sequencing reagents (Amersham Biosciences, Piscataway, NJ) with minor modifications of the manufacturer's recommendations and electrophoresed on a Li-Cor Global IR2 System slab-gel DNA sequencer through a 3.7% polyacrylamide gel matrix. Results were analyzed using E-seq 2.0 software (Li-Cor Biosciences, Lincoln, NE). In both cases, consensus sequences from multiple standard chromatogram format sequence (SCF) files were created using Sequencher 3.0 (Gene Codes Corp., Ann Arbor, MI) and aligned using the Clustal W algorithm in MacVector 7.2 (Accelrys Inc., San Diego CA) with default parameters. 12S and 16S rRNA alignments were adjusted by eye using secondary structure models of Ortí et al. (1996), Burk et al. (2002), and Wang and Lee (2002) following the methods of Kjer (1995). Ambiguous (unalignable) regions were excluded from further analyses to prevent loss of phylogenetic signal. Putative stem (paired) and loop/bulge (unpaired) regions were located in both data sets. Base pair complementarity in stem regions was confirmed by eye. Prior to alignment, all ND2 sequences were translated to ensure a single, continuous open reading frame. Adjustment of aligned ND2 consensus sequences was not necessary, as each sequence encoded a protein of the exact same length. The alignment of the ITS-1 region was performed on the basis of conserved sequence motifs using default pairwise gap opening and extension penalties in MacVector. Flanking transfer RNA (tRNA) locations within each mitochondrial alignment were located using tRNAscan SE (Lowe and Eddy, 1997). 18S and 5.8S gene regions were located in the nuclear ITS-1 alignment with the aid of ITS region sequences from *Auxis rochei* (Risso, 1810), Genbank accession AB193747.

Sequence features including nucleotide composition, site variability, and relative contribution by transitions (Ts) and transversions (Tv) were estimated using PAUP* 4.0b4 (Swofford,

1999). To infer sequence saturation, transitions and transversions were plotted against un-corrected sequence divergence (p-distance). Ts and Tv vs p-distance plots were generated for stem and loop partitions in the 12S and 16S data sets, for the first, second, and third positions in the protein-coding ND2 alignment, and for the ITS-1 data set overall.

To determine the validity of using a combined data set, a partition homogeneity test (Farris et al., 1994) was executed in PAUP* to test congruence between 12S, 16S, ND2, and ITS-1 data partitions. Incongruence length differences (ILDs) were explored with a heuristic search of 1000 replicates and 10 random sequence additions to assess data set congruence among data sets. DNA sequence alignments from 12S, 16S, ND2, and ITS-1 gene regions were analyzed separately and as two concatenated data sets (combined mitochondrial gene regions, and combined mitochondrial and nuclear gene regions) using three different inference methods: maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI).

Parsimony analyses were conducted using a heuristic search algorithm in PAUP* with tree-bisection-reconnection (TBR) branch swapping. Nonparametric bootstrapping (Felsenstein, 1985) using 1000 bootstrap pseudoreplicates with 1000 random sequence addition replicates were used to measure robustness of clade support for all data sets. Characters were unordered and equally weighted and gaps were considered missing information. Maximum likelihood analyses were performed using a heuristic search algorithm in PAUP* with base frequency, substitution rate and site variation parameters estimated using Modeltest 3.06 (Posada and Crandall, 1998). The choice of the most appropriate model was based on the Akaike Informa-tion Criterion (AIC). Searches were implemented in a heuristic maximum likelihood search with TBR branch swapping. Robustness of clade support was measured using 100 bootstrap pseudoreplicates with 10 random addition sequence replicates. Bayesian analyses were per-formed using MrBayes 3.1.1 (Ronquist and Huelsenbeck, 2003). For each individual gene re-gion data set, two Metropolis-coupled Markov chain Monte Carlo (MCMCMC) analyses were run for one million generations each, sampling every 1000 generations. Concatenated data sets were run for a total of four million generations sampling every 1000 generations. The GTR + I + G (general time reversible with a proportion of invariant sites and among site rate heterogeneity (gamma) parameter) model of character change was used in all analyses. In all cases, stationarity was reached after approximately 2000 generations. Data collected prior to this point were excluded in subsequent analyses to account for "burn in". Posterior prob-abilities were calculated as a measure of clade support with PAUP* using a 50% majority rule consensus. In all cases, phylogenetic trees were rooted with genetic data from *P. saltatrix*, *N. pectoralis*, and *C. armatus* (AP004444).

<div align="center">RESULTS</div>

Complete 12S rRNA, ND2, and ITS-1, and nearly complete 16S rRNA gene region sequences were isolated from all taxa, with the exception of *C. armatus*, which was downloaded from Genbank (AP004444). Complete 12S rRNA DNA sequences (Gen-Bank accession numbers FJ374786–FJ374798) ranged between 946 and 960 bp after removal of tRNA^{Phe} and tRNA^{Val}. Alignment was relatively straightforward following the proposed secondary structure models of Wang and Lee (2002). Two ambiguously aligned regions of length 20bp and 49bp were excluded from subsequent analyses, and were located in putative loop regions near the 5′ and 3′ end of the alignment, re-spectively. Nine-hundred and sixteen aligned bases were included in the final align-ment, of which 269 (29.4%) were variable, and 179 (19.5%) were parsimony informative (Table 2). Putative paired (stem) regions contained 464 nucleotide positions, of which 58 were parsimony informative. Putative unpaired (loop, bulge) regions contained 452 nucleotide positions, of which 121 were parsimony informative. A minor amount of transitional saturation was noted in unpaired regions, although not enough to warrant data exclusion or the use of an alternate character weighting scheme.

Nearly complete 16S rRNA gene region sequences were collected from all taxa. Sequences ranged between 1613 and 1660 bp after removal of tRNA[Val] and tRNA[Leu] fragments (GenBank accession numbers FJ374786–FJ374798). Alignment was relatively straightforward following the proposed secondary structure models of Ortí et al. (1996) and Burk et al. (2002). Two ambiguous regions of length 46bp and 44bp were excluded from subsequent analyses, and were located in putative loop regions in the central and 3′ end of the alignment, respectively. A total of 1599 aligned bases was included in the final alignment, of which 589 (36.8%) were variable, and 426 (26.6%) were parsimony informative (Table 2). Putative paired (stem) regions contained 718 nucleotide positions, of which 121 were parsimony informative. Putative unpaired (loop, bulge) regions contained 881 nucleotide positions, of which 305 were parsimony informative. No evidence of saturation was noted in paired regions. As was found in the 12S rRNA data set, a minor amount of transitional saturation was noted in unpaired regions, although not enough to warrant data exclusion or the use of an alternate weighting scheme.

ND2 gene region sequences from all taxa totaled 1047bp in length (GenBank accession numbers FJ374825–FJ374837). As there were no amino acid insertions or deletions in the data set, adjustment of the alignment was not necessary. Of the 1047 aligned bases 603 (57.6%) were variable, and 518 (49.5%) were parsimony informative (Table 2). One hundred and seventy-nine first position sites, 87 second position sites, and 337 third position sites were variable. While no evidence of saturation was found at first or second positions, severe transition saturation was noted in the third codon position. As a conservative measure, third position data were excluded from further analyses to reduce homoplasy. Of the remaining 698 aligned bases, 266 (38.1%) were variable and 196 (28.1%) were parsimony informative.

ITS-1 sequences were isolated from all taxa, with the exception of *C. armatus*; ITS-1 data from this species was not available. Two or three clones from each ITS-1 region were assayed to account for allelic variation. In all cases, only minor differences (base changes, insertions, deletions) existed between variants within an individual. The clone with the highest sequence quality (i.e., fewest within-sequence ambiguities) was chosen for phylogenetic analysis (GenBank accession numbers FJ374812–FJ374824). The final ITS alignment included complete ITS-1 region sequences which ranged between 416 and 725 bp, and partial flanking 18S and 5.8S region sequences which totaled 68 and 72bp, respectively. A total of 1010 aligned bases was included in the final alignment, of which 572 (56.6%) were variable, and 350 (34.7%) were parsimony informative (Table 2). A minor amount of transitional saturation was found, although not enough to warrant data exclusion or the use of an alternate weighting scheme.-

Phylogenetic Analyses.—Two concatenated gene region alignments were assembled and tested for data set congruence: a mitochondrial only (M-only) and mitochondrial plus nuclear (M + N) gene region alignment. Results of the partition homogeneity test indicate congruence between data partitions in the M-only alignment (P = 0.362). However, incongruence was found between data partitions in the M + N alignment (P = 0.001). Therefore, the primary result presented here (Fig. 1) is from the M-only data set, which had more informative characters than the ITS-1 (nuclear) data set. However, we also present results from a combined M + N analysis because this corroborates the M-only result despite the incongruity of the two data sets. ILD tests of data congruence have been shown to be inaccurate under certain

Table 2. Nucleotide sequence variation within the mitochondrial 12S rRNA, 16S rRNA, and ND2 gene regions and the nuclear ITS-1 region in the Echeneoidea.

| Region	Sequence length (bp)	No. variable sites	No. pars. inform. sites	Base frequencies				Base freq. homogeneity	Ts/Tv ratio
				A	C	G	T		
12S rRNA									
Stems	464	97 (20.9%)	58 (12.5%)	0.223	0.274	0.285	0.218	1.000	4.506
Loops	452	172 (38.1%)	121 (26.8%)	0.394	0.212	0.163	0.231	0.999	2.079
Overall	916	269 (29.4%)	179 (19.5%)	0.310	0.247	0.218	0.225	0.997	1.911
16S rRNA									
Stems	718	186 (25.9%)	121 (16.9%)	0.231	0.270	0.277	0.222	1.000	3.376
Loops	881	403 (45.7%)	305 (34.6%)	0.438	0.190	0.141	0.231	0.302	1.080
Overall	1,599	589 (36.8%)	426 (26.6%)	0.343	0.229	0.198	0.230	0.442	1.315
ND2									
1st position	349	179 (51.3%)	140 (40.1%)	0.303	0.303	0.187	0.207	0.819	2.506
2nd position	349	87 (24.9%)	56 (16.0%)	0.159	0.334	0.107	0.400	1.000	2.332
3rd position	349	337 (96.6%)	322 (92.3%)	0.355	0.328	0.073	0.244	0.000	1.416
Overall	1,047 (698)[a]	266 (38.1%)	196 (28.1%)	0.272	0.322	0.122	0.284	0.985	1.748
ITS-1									
Overall	1,010	572 (56.6%)	350 (34.7%)	0.172	0.303	0.336	0.189	0.010	0.994
Combined									
Mitochondrial only	3,213	1,124 (35.0%)	801 (24.9%)	0.308	0.252	0.197	0.243	0.137	1.881
Overall (mito + nuclear)	4,223	1,685 (39.9%)	1,164 (27.6%)	0.285	0.261	0.221	0.234	0.001	1.468

[a] Third position excluded
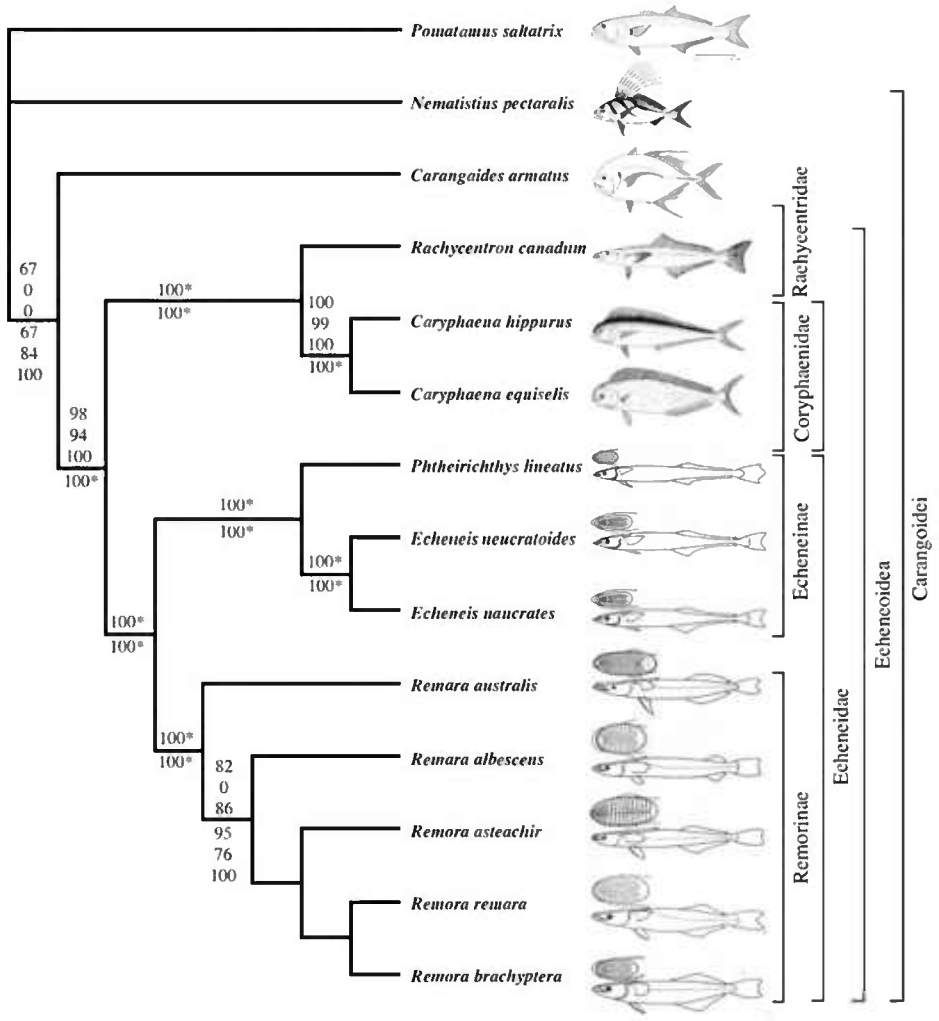
Figure 1. Phylogenetic estimates of the relationships within the Echenoidea, based upon both the M-only and M + N data sets using maximum parsimony. This tree was reflective of all analyses with the exception of the relationships within the Remorinae for which no support values are shown (see text for details). Upper support values refer to the M-only data set, while lower support values refer to the M + N data set. An * indicates that the support value was the same for MP, ML, and BI analyses. In cases where more than one value is present, values refer to MP, ML, and BI, respectively in vertical orientation.

conditions (Cunningham, 1997; Yoder et al., 2001). Furthermore, it has been argued that the combination of potentially incongruent (i.e., heterogeneous) data sets may increase phylogenetic accuracy (reviewed in Barker and Lutzoni, 2002). Data set characteristics of each concatenated alignment are noted in Table 2 and Table 3.

Results of the analysis of the M-only data set were largely similar regardless of the inference method used. In all cases, a monophyletic Echeneoidea was resolved with high bootstrap support (98 % MP; 94% ML) and high posterior probability (100% BI). Rachycentridae + Coryphaenidae formed a monophyletic sister group to the Echineidae (100%; all methods). Echeneidae was monophyletic (100%; all methods) and

Table 3. Summary statistics of parsimony and likelihood analyses of 12S rRNA, 16S rRNA, ND2, and ITS-1 DNA sequences in the Echeneoidea.

Region	Maximum parsimony					Maximum likelihood			
	Tree length	CI	RI	RC	HI	Evol. model	Prop. invar. sites (I)	Alpha shape parameter (G)	–ln Ta
12S rRNA	611	0.6007	0.4969	0.2985	0.3993	GTR + I + G	0.5481	0.7593	3,850.5
16S rRNA	1,334	0.6499	0.6311	0.4102	0.3501	GTR + I + G	0.3121	0.4836	7,884.9
ND2 1st and 2nd positions only	772	0.6010	0.5000	0.3347	0.3990	TIM + I + G	0.3822	0.6807	3,404.4
ITS-1	1,041	0.8204	0.7700	0.6317	0.1796	GTR + I + G	0.2675	4.0054	5,442.2
Combined data									
Mitochondrial only	2,458	0.6452	0.6220	0.4014	0.3548	GTR + I + G	0.4311	0.6863	15,270.2
Mitochondrial + nuclear	3,500	0.6906	0.6554	0.4526	0.3094	GTR + I + G	0.4329	1.1719	30,186.4

the subfamilies Echeneinae and Remorinae were both monophyletic (100%; all methods). The tree topology is shown in Figure 1. However, there were minor differences in the trees among the three optimality criteria. Parsimony analyses of the M-only combined data set resulted in a single most parsimonious tree of length 2458 (Consistency Index (CI) = 0.6452; Retention Index (RI) = 0.6220; Rescaled Consistency Index (RC) = 0.4014; Table 3). All nodes were supported by moderate (60%–85%) to strong (> 85%) bootstrap support values. A moderate level of support (67%) was estimated for an Echeneoidea + *C. armatus* clade. Within the Remorinae, *R. australis* was placed at the most basal position, whereas a *R. remora* + *R. brachyptera* clade occupies the most derived position. Maximum likelihood analyses produced a tree of somewhat different topology (–ln = 15270.2). Differences include the placement of *C. armatus* and the relationships within the Remorinae. *Nematistius pectoralis* and *C. armatus* form a monophyletic outgroup to the Echeneidae. Within the Remorinae clade, a polytomy of *R. albescens* + *R. remora* + *R. australis* + a weakly supported *R. osteochir* + *R. brachyptera* clade was resolved. Bayesian analyses resulted in of 3996 trees after removal of the burn-in. The topology of the consensus tree resembled that of the parsimony tree, except that *N. pectoralis* and *C. armatus* form a monophyletic outgroup to the Echeneidae, as was found in the likelihood tree. Posterior probabilities of clade support were strong (> 95%) at all nodes except within the Remorinae. As found in the parsimony tree, *R. australis* was the most basal member within the Remorinae.

Analyses of the M + N data set yielded congruent tree topologies to those generated by the M-only alignment. Once again, all three inference methods resulted in a monophyletic Echeneoidea with high bootstrap support (100% MP; 100% ML) and high posterior probability (100% BI). Rachycentridae + Coryphaenidae were resolved in all cases. Within a monophyletic Echeneidae, subfamilies Echeneinae and Remorinae were both monophyletic. Unlike the M-only data set, all methods resolved a Echeneoidea + *C. armatus* group with a moderate to strong level of bootstrap support (67% MP; 84% ML) and high posterior probability (100% BI). Parsimony analyses of these data resulted in a single most parsimonious tree of length 3500 (CI = 0.6906; RI = 0.6554; RC = 0.4526; Table 3). The tree topology recovered is identical to that found with the mitochondrial only data set, with slightly different bootstrap support values. Relationships within the Remorinae differed between trees generated using the three different inference methods. In all cases, however, *R. australis* was found at the most basal position and *R. brachyptera* was found within a clade at the most derived position. In all trees, *R. albescens* was placed within the Remorinae, although the exact placement varied among inference methods.

PHYLOGENETIC RELATIONSHIPS–INDIVIDUAL GENE REGIONS.—Individual gene phylogenies differed somewhat depending on the gene region examined and inference method used. 12S rRNA phylogenies agreed on the monophyly of the Echeneinae within a monophyletic, but polytomic Echeneidae clade. A monophyletic Rachycentridae + Coryphaenidae clade was resolved in all trees, with low to moderate bootstrap (63% MP; 53% ML), but high posterior probability (99% BI) support.

16S phylogenies closely resembled the results of the combined data analyses, with a few minor differences. A Rachycentridae + Coryphaenidae grouping was resolved with strong support in all cases (100%; all methods). Subfamilies Echeneinae and Remorinae were both monophyletic within a monophyletic Echeneidae. Parsimony analyses produced a topology identical to the one found using the M + N data set, ex-

cept *C. armatus* was not grouped with the monophyletic Echeneoidea clade. Likelihood analyses resolved a *Pomatomus saltatrix* + Echeneoidea clade, and an undefined relationship with the Echeneinae. Results of the Bayesian analyses were identical to those seen in the combined (mitochondrial only) data set, with the exception of the arrangement of species with the Remorinae.

Analyses of 698 bases of the ND2 gene yielded three trees of differing topology. In all cases, however, the Echeneoidea were monophyletic. In addition, a monophyletic Rachycentridae + Coryphaenidae clade was grouped sister to a monophyletic Echeneidae. Topologies within the Echeneidae differed significantly among inference methods. A monophyletic, polytomic and polyphyletic Remorinae was resolved using parsimony, likelihood and Bayesian inference, respectively. A monophyletic Echeneinae was resolved in all topologies.

Phylogenies based solely on analysis of ITS data were in agreement with most of the hypotheses generated using combined data. Each inference method resolved a monophyletic Echeneoidea. A monophyletic Rachycentridae + Coryphaenidae clade was resolved, sister to a monophyletic Echeneidae. The family Echeneidae was defined by monophyletic subfamilies Echeneinae and Remorinae. Relationships within the Remorinae differed among inference methods. Parsimony analyses placed *R. australis* in the basal position, and a clade containing *R. brachyptera* and *R. osteochir* at the most derived position. Likelihood analyses produced a polytomic arrangement. Bayesian analyses placed *R. brachyptera* in the basal position and a clade containing *R. australis* and *R. remora* at the most derived position.

Discussion

Hypotheses of the evolutionary relationships within the Echeneoidea were based on analyses of both concatenated mitochondrial DNA sequences and concatenated mitochondrial and nuclear DNA sequences. The hypotheses generated were largely consistent between M-only and M + N gene phylogenies, despite potential incongruence between mitochondrial and nuclear data sets. Overall, the phylogenetic hypotheses generated using the three different optimality criterion (parsimony, likelihood, Bayesian) were largely congruent. Nodal support at the superfamily, family, and subfamily levels was very high, as measured by nonparametric bootstrapping and posterior probabilities.

The taxonomic relationships presented here, observed in both concatenated M-only and M + N phylogenies, corroborate the family-level morphology-based hypotheses of Johnson (1984, 1993) and molecular hypotheses of Reed et al. (2002). These results contradict the morphology and behavior-based hypotheses of O'Toole (2002). Furthermore, the present results disagree with specific aspects of alpha level taxonomy hypothesized by O'Toole (2002). Below are the relevant findings, in reference to the six objectives of this study:

(1) In agreement with the work of Johnson (1984, 1993) and O'Toole (2002), the Echeneoidea were resolved as a monophyletic group. The Echeneidae, Coryphaenidae, and Rachycentridae were resolved together with strong support in all cases.

(2) The families Rachycentridae and Coryphaenidae form a monophyletic group. This hypothesis agrees with Johnson (1984, 1993), who cited a number of synapomorphies relating to neurocranial development, head spination, mandibular structure,

and epithelial cell composition in support of this relationship. This phylogeny also agrees with the hypotheses of Reed at al. (2002) and the work of Ditty et al. (1994) and Ditty and Shaw (1992).

(3) In agreement with the work of Lachner (1984) and O'Toole (2002), the Echeneidae form a monophyletic group.

(4) The subfamilies Echeneinae and Remorinae were both monophyletic, validating the subfamilial designations. These data contradict O'Toole (2002), who found the subfamily Echeneinae to be polyphyletic based on analyses of 138 putatively informative osteological characters and recommended the elimination of subfamilial designations.

(5) The genus *Remora* is paraphyletic based on the position of the monotypic genus *Remorina.* These results agree with O'Toole's (2002) findings. To amend this situation, O'Toole (2002) recommended that *R. albescens* be synonomized under the genus *Remora*, which yielded a monophyletic genus comprised of five members. Results of this study support that recommendation.

(6) The final objective, a clarification of the species-level relationships within the Echeneidae, was not fully achieved due to poor resolution within the Remorinae and potential discrepancies with the genus *Echeneis.*

In the analysis of combined data, two nodes were unresolved: the node defining the outgroups *N. pectoralis* and *C. armatus* and the node defining the Remorinae. *Carangoides armatus* was alternately grouped with either a monophyletic Echeneoidea or *N. pectoralis*. Freihofer (1978) united the five families Nematistiidae, Carangidae, Echeneidae, Coryphaenidae, and Rachycentridae based on two synapomorphies: an extension of the nasal canal surrounded by tubular ossifications and cycloid scales. Johnson (1984) and Smith-Vaniz (1984) further clarified the relationships with additional larval characters. The Carangidae were grouped with the Echeneoidea based on three characters: the lack of a bony stay posterior to the ultimate dorsal and anal pterygiophores, presence of two prenasal canal units, and a lamellar expansion of the coracoid. Parsimony, maximum likelihood, and Bayesian phylogenies of the M + N data set agree on the placement of *C. armatus* as a sister-taxon to the monophyletic echeneoid clade with a moderate (67% MP) to high (100% BI) level of support. Parsimony analyses of the combined M-only data set agrees with this relationship. On the contrary, likelihood and Bayesian analyses of these data resolve a highly supported monophyletic *N. pectoralis* + *C. armatus* clade. The phylogenetic relationships among these taxa were undefined by O'Toole (2002). In light of the previous, uncontested work of Johnson (1984) and Smith-Vaniz (1984), it seems highly likely that the grouping of [Nematistiidae + (Carangidae + Echeneoidea)] is the best-supported hypothesis.

The second unresolved aspect of the proposed echeneoid phylogeny involves the relationships within the Remorinae. Likelihood analyses of both the M-only and M+N data sets produced an unresolved relationship among the five member taxa. Parsimony and Bayesian analyses of both data sets agree on the placement of *R. australis* as the sister group to the remaining Remorinae. The relative positions of the remaining taxa varied among data sets and inference methods. As such, the relationships within this subfamily are still unresolved. To address this issue, future work should involve sampling a gene region that exhibits a higher rate of sequence evolution (e.g., the mitochondrial control region).

The level of genetic divergences between sister-species within the genus *Echeneis* is surprisingly low. Pairwise sequence divergence of 0.21% (12S rRNA), 0.25% (16S rRNA), 0.14% (ND2), and 0.66% (ITS) were estimated between the sharksucker (*E. naucrates*), and whitefin sharksucker *Echeneis neucratoides* Zuiew, 1786 specimens. Cursory analysis of these data suggests that these two specimens are in fact the same species, as this level of divergence is comparable to that found between conspecifics. For example, intraspecific sequence divergence between individuals of *R. brachyptera* exhibited the following ranges: 12S (0.32%–1.16%); 16S (0.18%–1.07%); ND2 (0.43%–1.00%) (data not shown).

There are two possible explanations for the observed levels of divergence between *Echeneis* species. First, one of the samples could have been misidentified. Alternatively, these two putative sister species could, in fact, comprise a species that simply demonstrates a wider range in character states and alternate color morphs than other echeneids. Debate currently exists as to whether *Echeneis* is monotypic. *Echeneis neucratoides* exhibits a restricted geographic distribution (western Atlantic only), whereas *E. naucrates* is found worldwide. These two species are nearly identical in external appearance and many of the putative characters used to differentiate the two species overlap. Only one putative sample of *E. neucratoides* was collected during this investigation. The specimen was identified by two separate researchers (K. Gray, B. B. Collette) using the keys of Lachner (1984) and Collette (2003), following standard identification procedures (including x-ray analysis of skeletal elements). On the basis of dorsal and anal fin coloration alone (as the other identifiable features fell within the range of both species), this specimen was identified as *E. neucratoides*. Although tree topology is unlikely to change, even if divergence values between these sister species were to increase to a level comparable to that seen between other remora species, these results highlight the need for inclusion of additional individuals to assess the validity of these two species. Moreover, multiple individuals of each member of the superfamily should be included, to ensure consistent and accurate phylogenetic interpretation.

Acknowledgments

Literature Cited

Barker, F. K. and F. M. Lutzoni. 2002. The utility of the incongruence length difference test. Syst. Biol. 51: 625–637.

Broughton, R. E. and J. R. Gold. 2000. Phylogenetic relationships in the North American cyprinid genus *Cyprinella* (Actinopterygii: Cyprinidae) based on sequences of the mitochondrial ND2 and ND4L genes. Copeia 2000: 1–10.

Burk, A., E. J. P. Douzery, and M. S. Springer. 2002. The secondary structure of mammalian mitochondrial 16S molecules: refinements based on a comparative phylogenetic approach. J. Mamm. Evol. 9: 225–252.

Chen, W. J., C. Bonillo, and G. Lecointre. 2003. Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. Mol. Phylogen. Evol. 26: 262–288.

Collette, B. B. 2003. Family Echeneidae. Pages 1414–1419 in K. E. Carpenter, ed. The living marine resources of the western Central Atlantic. FAO Species identification guide for fishery purposes and Amer. Soc. Ich. Herp. Spec. Publ. 5

Cunningham, C. W. 1997. Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. 32: 575–587.

Ditty, J. G. and R. F. Shaw. 1992. Larval development, distribution, and ecology of cobia *Rachycentron canadum* (family: Rachycentridae) in the northern Gulf of Mexico. Fish. Bull. 90: 668–677.

_____, R. F. Shaw, C. B. Grimes, and J. S. Cope. 1994. Larval development, distribution, and abundance of common dolphin, *Coryphaena hippurus*, and pompano dolphin, *C. equiselis* (family: Coryphaenidae), in the northern Gulf of Mexico. Fish. Bull. 92: 275–291.

Domanico, K. A., R. B. Phillips, and T. H. Oakle. 1997. Phylogenetic analysis of Pacific salmon (genus *Oncorhynchus*) using nuclear and mitochondrial DNA sequences Can. J. Fish. Aquat. Sci. 54: 1865–1872.

Farris, J. S., M. Kallersjo, A. G. Kluge, and C. Bult. 1994. Testing significance of incongruence. Cladistics 10: 315–319.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17: 368–376.

_____. 1983. Parsimony in systematics: biological and statistical issues. Annu. Rev. Ecol. Syst. 14: 313–333.

_____. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791.

Freihofer, W. C. 1978. Cranial nerves of a percoid fish, *Polycentrus schomburgkii* (family Nandidae), a contribution to the morphology and classification of the order Perciformes. Occas. Pap. Calif. Acad. Sci. No. 128: 1–78.

Johnson, G. D. 1984. Percoidei: development and relationships. Pages 464–498 in H. G. Moser et al., eds. Ontogeny and systematics of fishes. Spec. Publ. No.1, American Society of Ichthyologists and Herpetologists. Allen Press, Lawrence.

_____. 1993. Percomorpha phylogeny: progress and problems. Bull. Mar. Sci. 52: 3–28.

Johnson, K. 2003. A genetic analysis of the intraspecific relationships of tropical marine shorefishes common to Bermuda and the southeastern Atlantic coast of the United States. MSc. Thesis. The College of William and Mary: Virginia Institute of Marine Science. Gloucester Point, VA. 133 p.

Kjer, K. M. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. Mol. Phylogen. Evol. 4: 314–330.

Lachner, E. A. 1984. Family Echeneidiae. Pages 1329–1334 in P. J. P. Whitehead et al., eds. Fishes of the north-eastern Atlantic and Mediterranean. UNESCO, Paris.

Lowe, T. M. and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucl. Acids Res. 25: 955–964.

Mattern, M. Y. 2004. Molecular phylogeny of the Gasterosteidae: the importance of using multiple genes. Mol. Phylogen. Evol. 30: 366–377.

McDowell, J. R. 2002. Genetic stock structure of the sailfish, *Istiophorus platypterus*, based on nuclear and mitochondrial DNA. Ph.D. Diss. College of William and Mary. 231 p.

Musilová Z., O. Rícan, K. Janko, and J. Novák. 2008. Molecular phylogeny and biogeography of the Neotropical cichlid fish tribe Cichlasomatini (Teleostei: Cichlidae: Cichlasomatinae). Mol. Phylogen. Evol. 46: 659–672.

Near, T. J. C. and -H. C. Cheng. 2008. Phylogenetics of notothenioid fishes (Teleostei: Acanthomorpha): Inferences from mitochondrial and nuclear gene sequences. Mol. Phylogen. Evol. 47: 832–840.

Nelson, J. S. 2006. Fishes of the world. 4th ed. John Wiley and Sons, New York. 601 p.

Ortí, G., P. Petry, J. I. R. Porto, M. Jegu, and A. Meyer. 1996. Patterns of nucleotide change in mitochondrial ribosomal RNA genes and the phylogeny of piranhas. J. Mol. Evol. 42: 169–182.

O'Toole, B. 2002. Phylogeny of the species of the superfamily Echenoidea (Perciformes: Carangoidei: Echeneidae, Rachycentridae, and Coryphaenidae), with an interpretation of echeneid hitchhiking behaviour. Can. J. Zool. 80: 596–623.

Palumbi, S. R. 1996. Nucleic Acids II: The polymerase chain reaction. Pages 205–247 in D. M. Hillis, C. Moritz, and B. Mable, eds. Molecular systematics. Sinauer Associates, Sunderland.

Posada, D. and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817–818.

Reed, D. L., K. E. Carpenter, and M. J. deGravelle. 2002. Molecular systematics of the jacks (Perciformes: Carangidae) based on mitochondrial cytochrome b sequences using parsimony, likelihood and Bayesian approaches. Mol. Phylogen. Evol. 23: 513–524.

Ronquist, F. and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

Sambrook, J. and D. W. Russell. 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sanger, F., S. Nicklen, and A. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. 74: 5463–5467.

Smith-Vaniz, W. F. 1984. Percoidei: development and relationships. Pages 522–530 in W. J. Richards, D. M. Cohen, M. P. Fahay, A. M. Kendall, Jr., and S. L. Richardson, eds. Ontogeny and systematics of fishes. Spec. Publ.1, American Society of Ichthyologists and Herpetologists, Allen Press, Lawrence.

Sorenson, M. D., J. C. Ast, D. E. Dimcheff, T. Yuri, and D. P. Mindell. 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Mol. Phylogen. Evol. 12: 105–114.

Springer, V. G. and W. F. Smith-Vaniz. 2008. Supraneural and pterygiophore insertion patterns in carangid fishes, with description of a new Eocene carangid tribe +Paratrachinotini, and a survey of anterior anal-fin pterygiophore insertion patterns in Acanthomorpha. Bull. Biol. Soc. Wash. 16: 73 p.

Swofford, D. L. 1999. PAUP* Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland.

Thacker, C. E. 2003. Molecular phylogeny of the gobioid fishes (Teleostei: Perciformes: Gobioidei). Mol. Phylogen. Evol. 26: 354–368.

Wang, H. Y. and S. C. Lee. 2002. Secondary structure of mitochondrial 12S rRNA among fish and its phylogenetic implications. Mol. Biol. Evol. 19: 138–148.

Westneat, M. W. and M. E. Alfaro. 2005. Phylogenetic relationships and evolutionary history of the reef fish Labridae. Mol. Phylogen. Evol. 36: 370–390.

Yoder, A. D., J. A. Irwin, and B. Payseur. 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. Syst. Biol. 50: 408–424.

ADDRESSES: (K.N.G., J.R.M., J.E.G.) *Virginia Institute of Marine Science, College of William and Mary, P.O. Box 1346, Gloucester Point, Virginia 23062.* (B.B.C.) *National Marine Fisheries Service Systematics Laboratory, National Museum of Natural History MRC-0153, Washington, D.C. 20560.* CORRESPONDING AUTHOR: (J.E.G.) *E-mail: <graves@vims.edu>.*