

**MOLECULAR EVOLUTION IN THE CYTOCHROME B GENE OF
EPINEPHELINE FISHES (PERCOIDEI: SERRANIDAE)**

Joel L. Carlin
Department of Fisheries and Aquatic Sciences
University of Florida
7922 NW 71st Street
Gainesville FL 32653 USA
Fax: (352) 392-3672
Phone: (352) 392-9617
E-mail: joelcarlin@hotmail.com

Introduction

Groupers are members of the percoid family Serranidae, a "classificatory wastebasket" (Johnson, 1983) and possibly non-monophyletic family which lies within a suborder (Percoidei) and order (Perciformes) that each may not be monophyletic (Stokes, 1996). Grouper taxonomy contains several complexes of sympatric and parapatric species distinguished primarily by color pattern and combinations of overlapping meristic characters (e.g., the reticulated groupers of the Indo-West Pacific). Some species pairs differ only in color pattern and range while marked color pattern differences exist between distant populations within single widely distributed species (Heemstra and Randall, 1993). Resolution of such species complexes by molecular techniques addresses not only evolutionary history but allows for more informed management in multi-species fisheries.

I present the preliminary analyses of the mitochondrial gene cytochrome *b* (*cyt b*) from nearly a third of the subfamily Epinephelinae (*sensu* Heemstra and Randall, 1993). Because epinepheline fishes have such a broad distribution and historically plastic taxonomy, I chose to maximize the intraspecific collection whenever possible. Therefore the survey considered each specimen individually rather than using a consensus sequence (except for identical haplotypes shared among specimens), allowing *cyt b* to be evaluated for phylogenetic utility from the interfamily to the intraspecific level. A reasonable expectation is that strong phylogenetic signal exists with minimal homoplasy at some but not all levels of analysis (Allegrucci et al., 1999), and that the use of separate haplotypes from

different specimens may better represent the range of genetic variability than by limited intraspecific sampling.

When considering a large number of taxa ($m > 50$) from a modest dataset ($n < 1000$ nucleotides), the relationship between model sophistication / tree search method and the probability of recovering a correct topology is less stringent (Nei and Kumar, 2000). What model of DNA substitution best describes the evolution of *cyt b* in groupers, and does this model allow for the recovery of a correct topology? How are alternate tree search methods affected in such a dataset? I addressed these questions by comparing topologies from multiple tree-building methods with the acknowledgement that without a known phylogeny, concordance among trees reflects precision, not accuracy. Therefore my objectives were to examine the patterns of substitution present, the topological differences among tree construction methods, and differences between the *cyt b* gene tree and other hypotheses.

Materials and Methods

An approximately 750 base pair (bp) fragment of the mitochondrial *cyt b* gene was amplified by procedures outlined in Carlin et al. (in review). All samples were sequenced at least twice with the amplification primers by the author or by the University of Florida DNA Sequencing Core or the Nevada Genomics Center. Several haplotypes each were chosen to represent the range of both geographic distance and genetic divergence observed across the species range of *Epinephelus adscensionis* and *Rypticus saponaceus* (Carlin et al., in review). Sequence polymorphism and substitution model selection were conducted in Paup*β5.10 (Swofford, 2000) and Modeltest 2.0 (Posada and Crandall, 1998) as described by Cunningham et al. (1998) and Posada and Crandall (1998).

Pairwise comparisons among taxa were analyzed at four levels of *a priori* taxonomic hierarchy: among all sequences within a species (intraspecific), among single randomly chosen sequences from each species within a genus (interspecific), between species in different genera only (intergeneric), and over all taxa. Because a 16S rRNA gene tree of some epinephelins indicated *Epinephelus* was polyphyletic relative to *Mycteroperca* (Craig et al., 2001), some analyses were performed on a reduced dataset including *Epinephelus* and *Mycteroperca* sequences only.

Four topologies each were generated for the whole dataset (all serranids + putative outgroups) and reduced dataset (tribe Epinephelini only). A neighbor-

joining (NJ) distance phylogram was constructed under the substitution model preferred by the likelihood ratio test. The topology of the preferred-method NJ tree was compared to a NJ tree constructed by simple proportional (p) distance (Tajima and Takezaki, 1994; Russo et al., 1996; Nei and Kumar, 2000). Node precision was assessed by 500 iterations each of bootstrap resampling (e.g., resampling with replacement of nucleotides) and jackknife resampling (e.g., resampling with replacement of taxa). Maximum parsimony (MP) gene trees were also constructed, but not by searching all available treespace. To compare tree search efficiencies and topologies, 50 iterations of heuristic searches by random taxon addition were used, with each addition iteration followed by either nearest neighbor interchange swapping or the more exhaustive tree bisection-reconnection branch swapping protocol (Swofford and Begle, 1993). Additional trees examined alternate ti:tv weighting and outgroup designations.

Results

A total of 99 sequences were examined, including 45 identified epinepheline species from ten genera (out of a possible 159 spp. in 15 genera listed by Heemstra and Randall, 1993) and 11 species from other subfamilies or families. Four epinepheline genera (7 species) were not available for analysis. Sequences were also obtained from the serranid subfamilies Serraninae (6 species of *Serranus*) and Grammistinae (*Aporops* spp. A and *Rypticus saponaceus*). Four sequences obtained from GenBank were used to represent the perciform families Haemulidae, Nemipteridae, Percichthyidae and Polyprionidae. Six haplotypes were identified only to genus: *Epinephelus* spp. A (GenBank Accession AF143193; Allegrucci et al., 1999) and *Epinephelus* spp. B (collected by the author from commercial markets in Hong Kong), 1 *Mycteroperca* spp. (mid-Atlantic islands of the Azores, Portugal), 2 specimens of *Serranus* spp. A (Gulf of Guinea, East Atlantic) and 1 *Aporops* spp. (KU 804).

A total of 134 specimens yielded 84 unique haplotypes in 607 bp from the cytochrome *b* mitochondrial gene. The alignment contained 314 constant and 265 parsimony-informative characters, and no indels or stop codons. The alignment was poor in guanines (23.4% A, 30.3% C, 15.8% G; $X^2 = 134.7$, d.f.=243, $P=1.00$), especially when only third codon sites were considered (42.9% C, 5.1% G; $X^2 = 364.7$, d.f.=243, $P<0.01$). The number of transitions (ti) did not increase linearly with the transversions (tv) in plots of absolute numbers of substitutions for most pairwise comparisons, indicating multiple substitutions per site, or saturation. Instead, pairwise comparisons with >20 tv from all codon positions have a wide range (approx. 50-90) of ti, encompassing 49.6% of all

interspecific pairwise comparisons. The greatest saturation effect was evident at the third codon position. Substitution prevalence varied by codon position (16:5:100 and 1:0:20 for first, second, and third positions in interspecific and intraspecific comparisons, respectively) and by steric type (73:54 interspecific ti:tv; 10:1 intraspecific ti:tv).

Two or more specimens were sequenced for 24 species (42.9% of species surveyed). Mean pairwise divergence between interspecific haplotypes within genera was $d = 18.36\%$ (range: 2.4-10.54%). The greatest differentiation within a genus occurred among *Cephalopholis* sequences ($d = 16.4\%$). The least interhaplotype divergence estimates occurred at *Plectropomus leopardus* vs. *P. maculatus* (9 ti, 0 tv), *Epinephelus niveatus* vs. *E. niphobles* (10 ti, 0 tv), *Mycteroperca bonaci* vs. *M. rosacea* (8 ti, 3 tv) and *E. chlorostigma* vs. *E. cyanopodus* (47 ti, 4 tv). Multiple haplotypes within a single species were monophyletic for each species (>98% support after 500 bootstrap and 500 jackknife iterations). However, strong support was also recorded for deep divergences within monophyletic *Epinephelus adscensionis* ($d = 4.3\%$), *Rypticus saponaceus* ($d = 4.4\%$), and *Variola louti* ($d = 10.4\%$). Of the 24 species with >1 specimen, haplotypes were shared across oceans in *Cephalopholis cruentata* (Bahamas and São Tomé), *Epinephelus adscensionis* (Bahamas and São Tomé), *Rypticus saponaceus* (São Tomé and Cape Verde; Ascension and Florida), and *Variola louti* (Australia and Seychelles; Australia and Philippines). In addition, two specimens of *M. phenax* from the Gulf of Mexico shared a haplotype with the unidentified *M. spp. A* from the Azores.

The substitution model preferred by likelihood ratio tests of all serranids in the dataset (one randomly chosen sequence per species) was the general time reversible (GTR) model (Nei and Kumar, 2000) with proportion of invariable sites = 0.393 and corrected by a γ distribution ($\alpha=0.869$). Distance trees constructed using p-distances and GTR- γ had very different topologies. However, comparisons of NJ trees after 500 bootstrap and jackknife iterations reduced many of the disputed nodes to unresolved polytomies while retaining several groups of sequences. Five nodes were present in the p-distance NJ tree that were not present under a GTR- γ substitution model. In all five instances the p-distance nodes were poorly supported (51-65% bootstrap support) and were members of a larger polytomy. Retention of these nodes reflects the generally larger bootstrap and jackknife support in the p-distance NJ tree relative to the GTR- γ NJ tree, with the greatest differences seen at more inclusive nodes.

Outside the *Epinephelus*+*Mycteroperca* phylogroup lay an unresolved polytomy of the remaining genera. All *Plectropomus* sequences were consistently unified. The p-distance NJ tree could also unify all *Serranus* and almost all *Cephalopholis* (except *C. nigri*). Arrangements of genera outside the *Epinephelus*+*Mycteroperca* clade varied only slightly. *Cephalopholis* was ancestral to the rest of the Epinephelini by p-distance but not by the GTR- γ model. Both models joined *Saloptia*, *Plectropomus* and the nemipterid sequence and the haemulid sequence unified with the grammistines.

Maximum parsimony cladograms were generated by two different heuristic search protocols (TBR and NNI branch swapping) and under two different weighting schemes (equal and 1:5 tv:ti weighting) for the entire dataset and for the Epinephelini only. Search time using NNI branch swapping was 10-14 times faster than by TBR branch swapping and resulted in slight differences in topology. Not surprisingly, distance phylogroups well-supported by bootstrap and jackknife resampling were also well-supported clades in most or all parsimony tree searches. Six node changes were present among parsimony trees (Figures 1, 2). The *E. bleekeri*/*E. chlorostigma* phylogroup was derived relative to *E. miliaris* and then to *E. quoyanus* in TBR and 5:1 NNI searches but not in 2:1 NNI searches. The sister taxon of *E. fasciatus* was *E. tauvina* by NNI searches and *E. caeruleopunctatus* by TBR searches. *E. itajara* and *E. polyphekadion* were derived relative to *E. fuscoguttatus*; these three species were unresolved in the 5:1 NNI search. *Mycteroperca fusca* was ancestral to *E. marginatus* in the 5:1 TBR search but otherwise all tree searches placed *E. marginatus* as an ancestor to a monophyletic *Mycteroperca*. Lastly, *Amyperodon leucogrammicus* had an uncertain position, being sister taxon of the *Epinephelus* spp. A (Allegrucci et al., 1999) sequence by 2:1 NNI, the sister taxon of *E. fuscoguttatus* in TBR, and unresolved by in 5:1 NNI. *Serranus phoebe* and *S. tabacarius* were sister taxa in 2:1 weightings by both searches, but otherwise were part of an unresolved polytomy within the monophyletic genus *Serranus*.

A strict consensus of 66 most parsimonious MP trees as generated by TBR branch swapping is presented in Figures 2 and 3. Dashed lines represent connections not presented by one or more of the three other tree building methods (the node collapses to a polytomy). The serranid parsimony tree was rooted to a polytomy of sequences from the Haemulidae and Nemipteridae. The subfamily Epinephelinae was polyphyletic; serranine and grammistine serranids as well as several presumptive outgroup sequences were more derived than the epinephelines *Saloptia*, *Plectropomus*, and *Variola*. There was also no support for a monophyletic *Epinephelus*, the most speciose epinepheline genus.

Figure 1. Derived groupers. Strict consensus MP cladogram by heuristic TBR swapping (CI = 0.177, RI = 0.631), with bootstraps. Dashed lines are connections not observed by distance or NNI swapping.

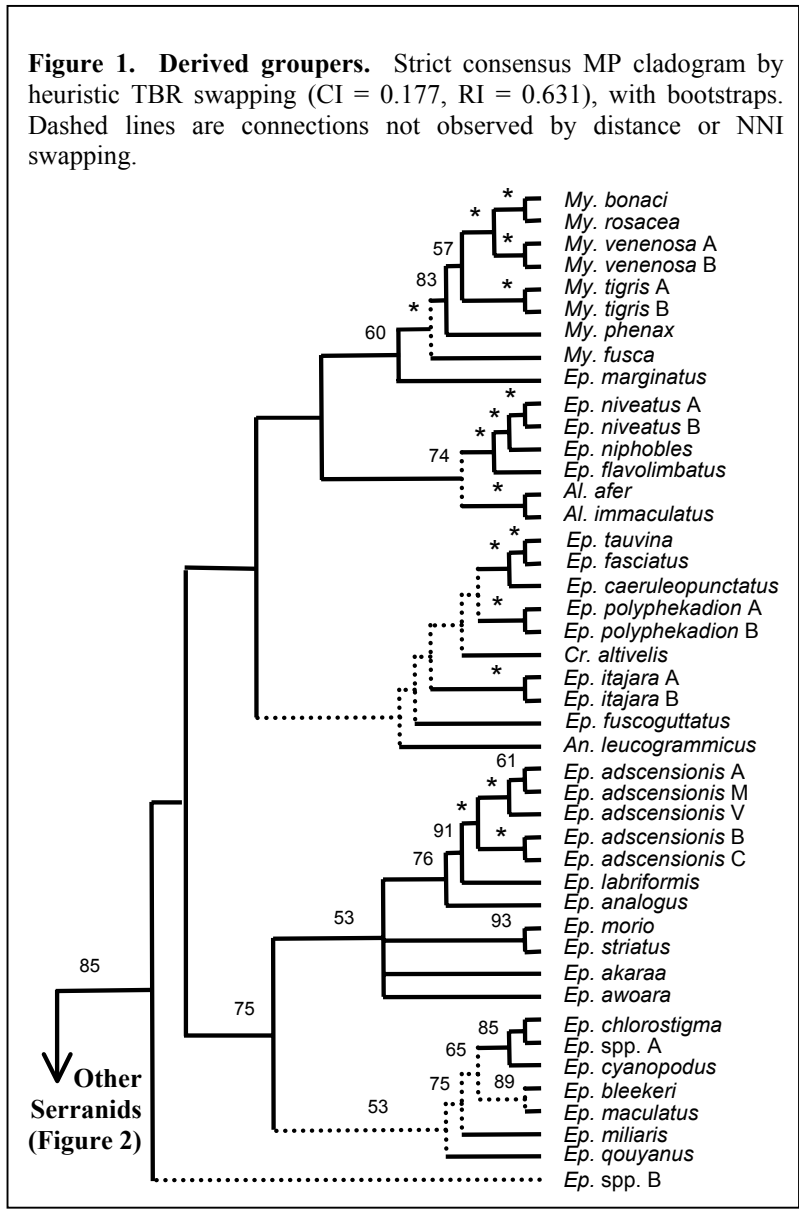
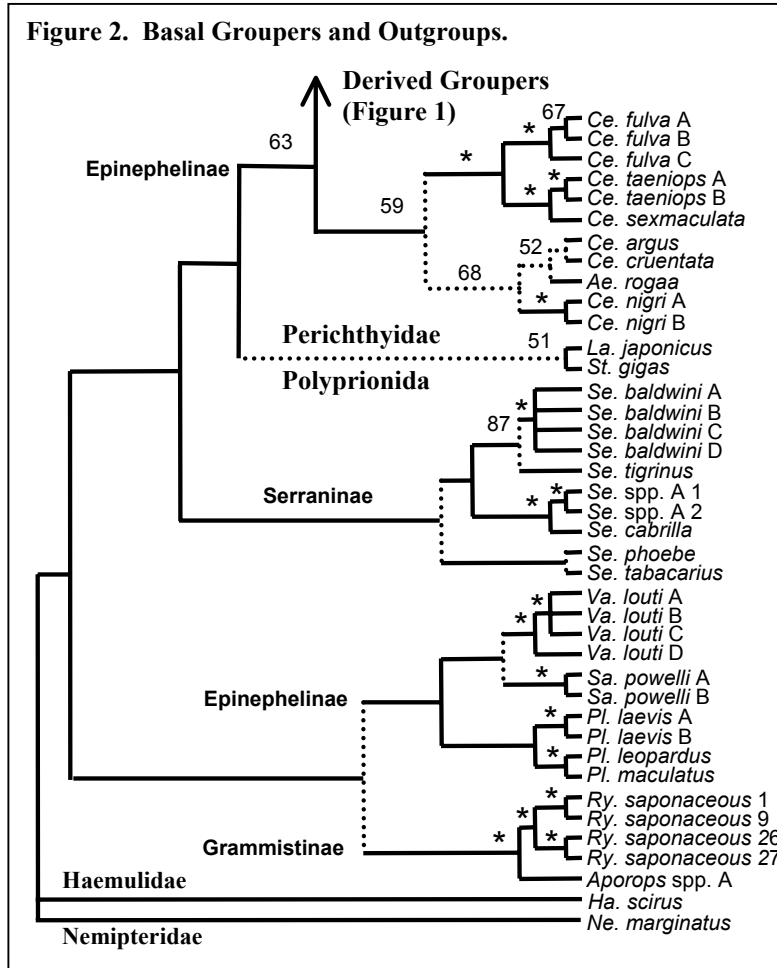


Figure 2. Basal Groupers and Outgroups.



A moderately supported near-polytomy (the *Epinephelus*+*Mycteroperca* clade) was found consistently in both NJ and MP trees and included *Epinephelus* spp. and *Alphestes*, *Anyperodon*, *Cromileptes*, and *Mycteroperca*. Groups outside the Epinephelini had poor resampling support by parsimony and were in near-unresolved distance polytomy of genera. The epinepheline genera *Cephalopholis* and *Plectropomus* and the serranine *Serranus* had poor support in the parsimony cladograms. The monotypic genus *Aethaloperca* was within an otherwise monophyletic *Cephalopholis*; this clade was the closest ancestor to the Epinephelini clade in all parsimony trees. The grammistines (*Aporops* + 4 *R. saponaceous* haplotypes) formed a strongly supported monophyletic group. Otherwise, eight groups of species within genera could be described as monophyletic with 70-100% bootstrap and jackknife support.

Discussion

At this locus, the range of variation over all positions was sufficient to create clades of species present in both MP and NJ trees, well-supported by bootstrap and jackknife resampling. The p-distance NJ tree showed best agreement with downweighted transition MP cladograms, especially nearer to terminal taxa. Divergences between terminal taxa indicate potential rate variation among the epinephelins and other percoid fishes. Sister taxa reported from the East Pacific and West Atlantic, respectively, are separated by two levels of genetic divergence: [*E. labriformis* + *E. adscensionis*] and [*A. immaculatus* + *A. afer*] have pairwise divergences of $d = 0.095 - 0.100$ whereas [*E. niveatus* + *E. flavolimbatus*] and [*M. rosacea* + *M. bonaci*] have a tenfold lower level of divergence ($d = 0.016-0.018$). Therefore rate variation will be addressed critically in future analyses.

These data demonstrate that species represented by multiple specimens can reveal hidden polyphyletic taxa, as seen in the divergent lineages identified within *E. adscensionis*, *R. saponaceous*, and *V. louti*. An additional species, *E. marginatus*, has divergent forms living in partial sympatry in the western Mediterranean ($d = 0.158$). From these results it seems likely that many species of serranids could have cryptic phylogroups. Gilles et al. (2000) interpreted the two *E. marginatus* as separate species, and the deep splits seen here are similar in depth to most interspecies pairwise comparisons in marine fishes (Johns and Avise, 1998). Yet the taxonomic significance of such genetic breaks is not necessarily certain (Palumbi, 1994; Nichols, 2001). Whether or not the divergent mtDNA lineages observed here are separate species, these data

demonstrate the importance of extensive geographic sampling to reveal a complete evolutionary history.

In order to examine the effects of incomplete taxon sampling, a jackknife resampling procedure was used to estimate the influence of individual sequences upon character variability and ordination. The greatest jackknife support (99-100%) was consistently present for species with multiple haplotypes. The node joining the alternate lineage groups in *E. adscensionis* is monophyletic with 100% jackknife support, as is the [*E. adscensionis* B + C] node (Figure 1). Yet the node containing haplotypes A, M, and V has only moderate jackknife support (70%) and indeed it is this A clade which contains 23 of the 28 described haplotypes (Carlin et al., in review). Jackknife support at the node uniting two *Alphestes* species (of three described) was strong (94%) but not total, implying the potential existence of an unsampled taxon between the two sampled *Alphestes*. The inclusion of additional species is both desirable and in progress, but the preliminary *cyt b* gene tree can be compared to other trees.

When this study is compared with to the phylogeny presented by Smith (1971) and the 16S gene tree of Craig et al. (2001) are considered, many similarities are notable among the relatively few taxa present in all trees. As Smith (1971) hypothesized, *Alphestes* is a descendant from an *Epinephelus* lineage by both *cyt b* and 16S. The genus *Mycteroperca* was within a polyphyletic *Epinephelus* (as in 16S) and not ancestral to *Epinephelus* as suggested by Smith (1971). The genus *Cephalopholis* was ancestral to *Epinephelus*, as in 16S but contradicting its derived status by Smith (1971). Grammistines and *Plectropomus* were sister taxa at both 16S and *cyt b*, which may revise the hypothesized position of a monophyletic Epinephelinae as the ancestor of the Grammistinae (Johnson, 1983). Removal of *Plectropomus* from the Epinephelinae could resolve this conflict, as well as better explain larval apomorphies (Leis, 1986). Overall, the deeper nodes were consistent between 16S and *cyt b*, but the level of concordance is difficult to test due to having only 15 species in common.

The congruence between the preliminary gene tree presented here and other phylogenetic hypotheses probably reflects real evolutionary history, especially regarding the paraphyly and polyphyly that is apparently rampant in serranids. The use of multiple specimens per species allows the documentation of the geographic range of shared haplotypes and cryptic phylogroups, allowing a more accurate measure of extant species diversity and more precise estimation of speciation rates. The p-distance tree provided more resolved nodes with

greater support than a distance tree constructed under conditions recommended by likelihood ratio tests. Since TBR branch swapping examined a greater portion of all treespace, so this method is recommended over NNI searches. However, investigators wishing for a quick approximation of major evolutionary patterns may well consider the NNI search protocol, since computation by this method was 14 times faster than TBR and resulted in mostly minor topological differences. The application of these results to future analytical methods will result in refined grouper taxonomy and hopefully a more complete understanding of evolution of marine taxa.

References

- Allegrucci, G., A. Caccone and V. Sbondoni. 1999. Cytochrome *b* sequence divergence in the European sea bass (*Dicentrarchus labrax*) and phylogenetic relationships among some Perciformes species. *J. Zool. Syst. Evol. Res.* 37:149-156
- Avise, J.C. 2000. *Phylogeography: The history and formation of species.* Harvard University Press, Cambridge, Massachusetts
- Carlin, J.L., D.R. Robertson and B.W. Bowen. In review. Ancient divergences and recent connections in two tropical Atlantic reef fishes *Epinephelus adscensionis* and *Rypticus saponaceous* (Percoidae: Serranidae). *Mar. Biol.*
- Coleman, F.C., C.C. Koenig, G.R. Huntsman, J.A. Musick, A.M. Eklund, J.C. McGovern, R.W. Chapman, G.R. Sedberry and C.B. Grimes. 2001. Long-lived reef fishes: the grouper-snapper complex. *Fisheries* 25:14-21
- Craig, M.T., D.J. Pondella, J.P.C. Franck and J.C. Hafner. 2001. On the status of the serranid fish genus *Epinephelus*: Evidence for paraphyly based on 16S rDNA sequence. *Mol. Phylo. Evol.* 19:121-130
- Cunningham, C.W., H. Zhu and D.M. Hillis. 1998. Best-fit maximum likelihood models for phylogenetic inference: empirical tests with known phylogenies. *Evolution* 52: 978-987
- Gilles, A., A. Miquelis, J.P. Quignard and E. Faure. 2000. Molecular phylogeography of western Mediterranean dusky grouper *Epinephelus marginatus*. *Compt. Rendus Acad. Sci.* III 323:195-205

- Hasegawa, M., H. Kishino and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160-174
- Heemstra, P.C. and J.E. Randall. 1993. Groupers of the world (Family Serranidae, subfamily Epinephelinae) : an annotated and illustrated catalogue of the grouper, rockcod, hind, coral grouper and lyretail species known to date. FAO, Rome Italy
- Johns, G.C. and J.C. Avise. 1998. Tests for ancient species flocks based on molecular phylogenetic appraisals of *Sebastes* rockfishes and other marine fishes. *Evolution* 52:1135-1146
- Johnson, G.D. 1983. *Nippon spinosus*: A primitive epinepheline serranid, with comments on the monophyly and intrarelationships of the Serranidae. *Copeia* 1983:777-787
- Leis, J.M. 1986. Larval development in four species of Indo-Pacific coral trout *Plectropomus* (Pisces: Serranidae: Epinephelinae) with an analysis of the relationships of the genus. *Bull. Mar. Sci.* 38:525-552
- Nei, M. and S. Kumar. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York, New York
- Nichols, R. 2001. Gene trees and species trees are not the same. *TREE* 16:358-363
- Palumbi, S.R. 1994. Genetic divergence, reproductive isolation, and marine speciation. *Ann. Rev. Ecol. Syst.* 25:547-572
- Posada, D. and K. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817-818
- Russo, C.A.M., N. Takezaki and M. Nei. 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Mol. Biol. Evol.* 13: 525-536
- Smith, C.L. 1971. A revision of the American Groupers: *Epinephelus* and allied genera. *Bulletin of the American Museum of Natural History* 146:1-241

- Stokes, R.T. 1996. Phylogenetic analysis of western Atlantic grouper (Epinephelinae) and other serranids (Teleostei: Serranidae) by peptide mapping of homologs of LDH-A4. M.S. Thesis. University of South Florida, Orlando Florida
- Swofford, D.L. 2000. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Massachusetts
- Swofford, D.L. and D.P. Begle. 1993. PAUP: Phylogenetic Analysis Using Parsimony, ver 3.1, user's manual. Illinois Natural History Survey, Champaign Illinois
- Tajima, F. and N. Takezaki. 1994. Estimation of evolutionary distance for reconstructing molecular phylogenetic trees. *Mol. Biol. Evol.* 11:278-286
- Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10:512-526.
- Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. *TREE* 11:367-372

Acknowledgements

Many persons contributed time and samples, but primary acknowledgements are given to thank A.M. Ecklund, K. Moots, K. Rhodes, D.R. Robertson, D. Weaver, and Y. Sadovy. A.C. Bentley, S.D. Gardieff, and R.H. Robins assisted with specimens from the Florida Museum of Natural History and the Univ. of Kansas Museum of Natural History. The author was funded by the AFS Genetics Section, AFS Florida Chapter, International Women's Fishing Association, PADI Project AWARE, the PADI Foundation and the University of Florida Department of Fisheries and Aquatic Sciences.